Project Partners

Swideas

1506
UNIVERSITÀ
DEGLI STUDI
DI URBINO
CARLO BO

Lai-momo

SIMORA
RAZVOJNA AGENCIJA
SISAČKO MOSLAVAČKE ŽUPANIJE

eurada

INSPIRING REVOLUTIONARY EDUCATIONAL CREDENTIALS

MODULE 13

One Block for Educational Credentials (OBEC)
2020-1-SE01-KA204-077803

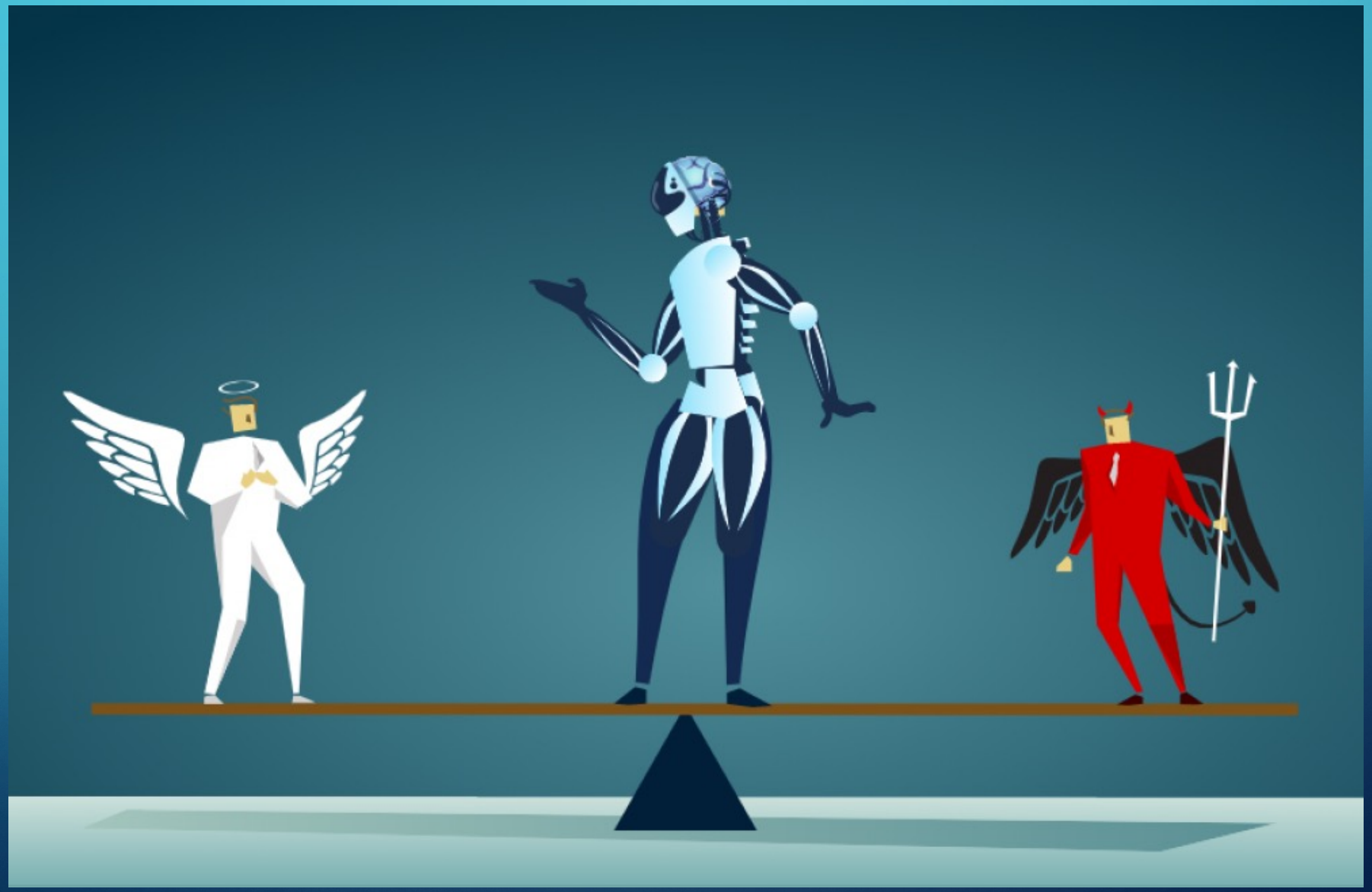
Co-funded by the
Erasmus+ Programme
of the European Union

PROBLEMI ETICI DELL'INTELLIGENZA ARTIFICIALE

PROF. MIRKO TAGLIAFERRI

BEC

MODULO 1: ETICA DELL'INTELLIGENZA ARTIFICIALE



CONTENUTI DELLA LEZIONE

1. Introduzione
   - Definizioni chiave e metodologia.
2. Dibattiti principali
   - Temi chiave dell'etica dell'intelligenza artificiale.
3. Conclusione:
   - Come procedere?

INTRODUZIONE

Quando si parla di etica della tecnologia esistono problemi e pseudo-problemi. Riuscire ad operare una corretta catalogazione è il primo passo per trovare una risposta efficace alle sfide etiche legate alle tecnologie.

PROBLEMI DATATI E/O INESISTENTI

Quando nacquero i primi treni a vapore, si credeva che il corpo delle donne non fosse adeguato a spostarsi a velocità superiori a 80km/h e che il loro utero si sarebbe staccato dal corpo.

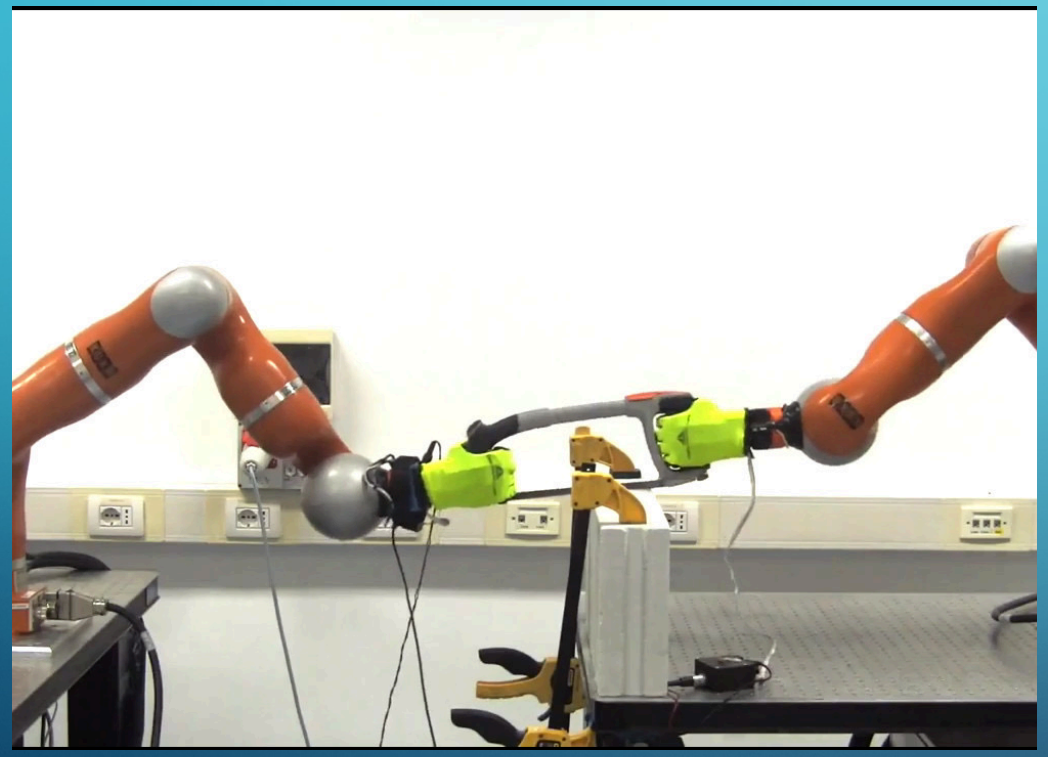
PROBLEMI SOVRASTIMATI

Quando la televisione entrò a far parte della grande distribuzione si pensava che essa avrebbe sancito il fallimento dell'intera industria cinematografica.

PROBLEMI REALI MA POCO INCISIVI

L’introduzione delle fotocamere digitali ha portato al lento declino dell’industria della produzione delle pellicole fotografiche.

PROBLEMI REALI ED ESTREMAMENTE RILEVANTI

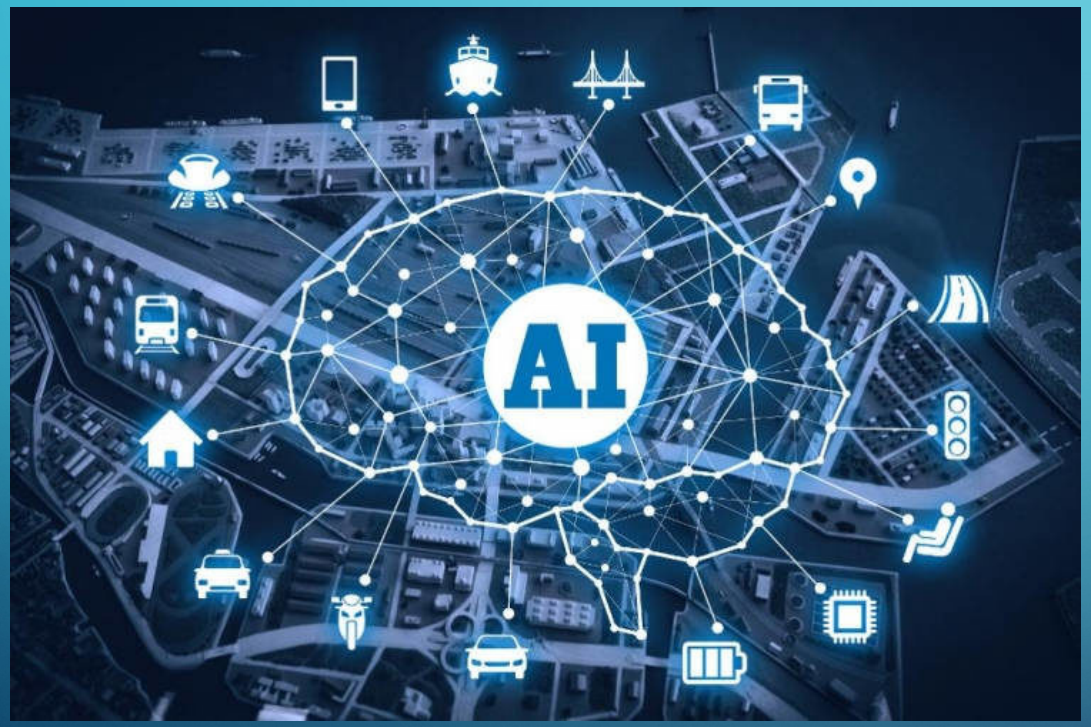


La robotica sostituirà gradualmente tutti gli impieghi a bassa specializzazione. Questo comporterà la perdita di innumerevoli posti di lavoro.

PRIMO PASSO

- Il primo passo quando si analizza un problema etico legato alla tecnologia è categorizzare bene il problema. Occorre comprendere se esso è, effettivamente, un problema. Occorre poi dare le giuste dimensioni a tale problema. Solo allora si potrà valutare se esiste una possibile soluzione a tale problema, trovando le giuste norme e i corretti sistemi concettuali per trattarlo.

INTRODUZIONE: AI



Artificial Intelligence (AI): qualunque sistema computazionale artificiale che mostra segni di comportamento intelligente. Un comportamento è definito come intelligente se tale comportamento è complesso e indirizzato verso l'ottenimento di un obiettivo.

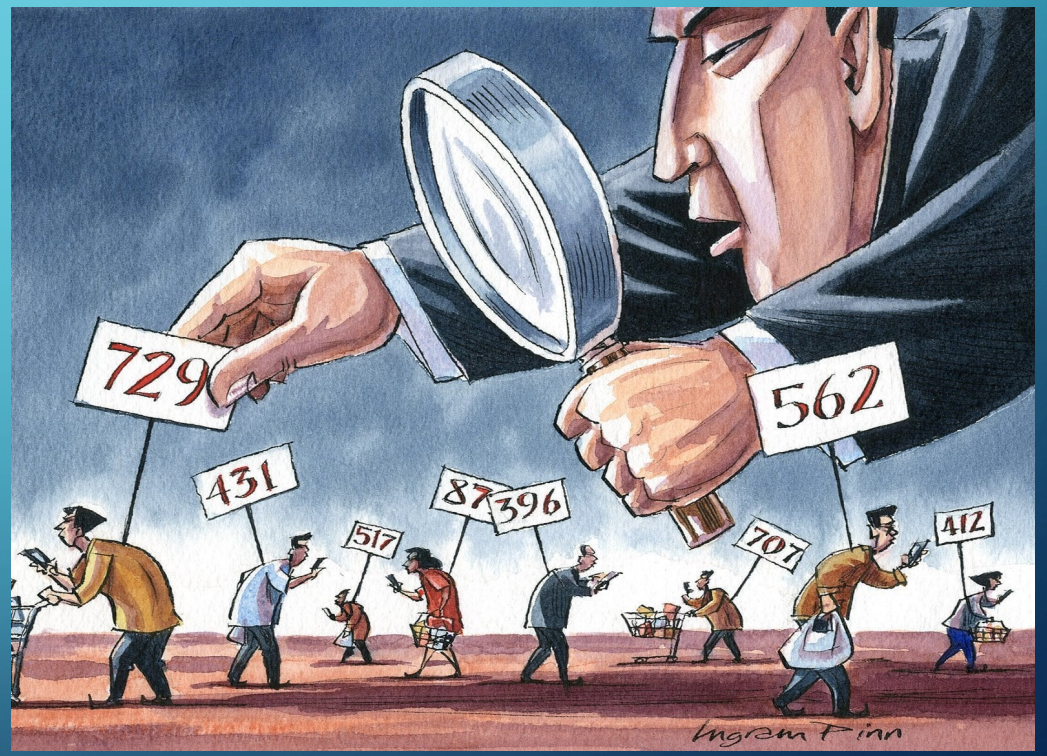
CHIARIRE LA DEFINIZIONE

- Non saranno importanti le specifiche di come una AI arriva a mostrare le caratteristiche che possiede.
- Per questo modulo conteranno come AI sia sistemi di manipolazione simbolica classici (e.g., dimostratori automatici) sia sofisticati meccanismi di apprendimento automatico (e.g., reti neurali).
- Inoltre, per i problemi analizzati, si prenderanno ad esempio sia AI completamente basate su software che comprendenti un corpo fisico.

DIBATTITI PRINCIPALI

1. Privacy e Sorveglianza
2. Manipolazione del Comportamento
3. Opacità dei Sistemi IA
4. Bias delle IA

1. PRIVACY E SORVEGLIANZA



SORVEGLIANZA: IL SISTEMA DEL CREDITO SOCIALE

https://www.youtube.com/watch?v=u2evek-JPAM&t=79s

PRIVACY

- Controllare chi colleziona i nostri dati e chi abbia accesso ad essi è sempre più difficile nel mondo digitale. Non solo: le moderne tecniche di analisi di tali dati permettono a sistemi digitalizzati di avere un'immagine estremamente accurata di noi come esseri umani.

BIG DATA

- Queste rappresentazioni virtuali del nostro essere hanno un enorme impatto, poiché attraverso di esse ci vengono forniti dei servizi specifici.
- Non solo, tali immagini rimangono indelebili e costantemente presenti nelle reti, *in primis* Internet.

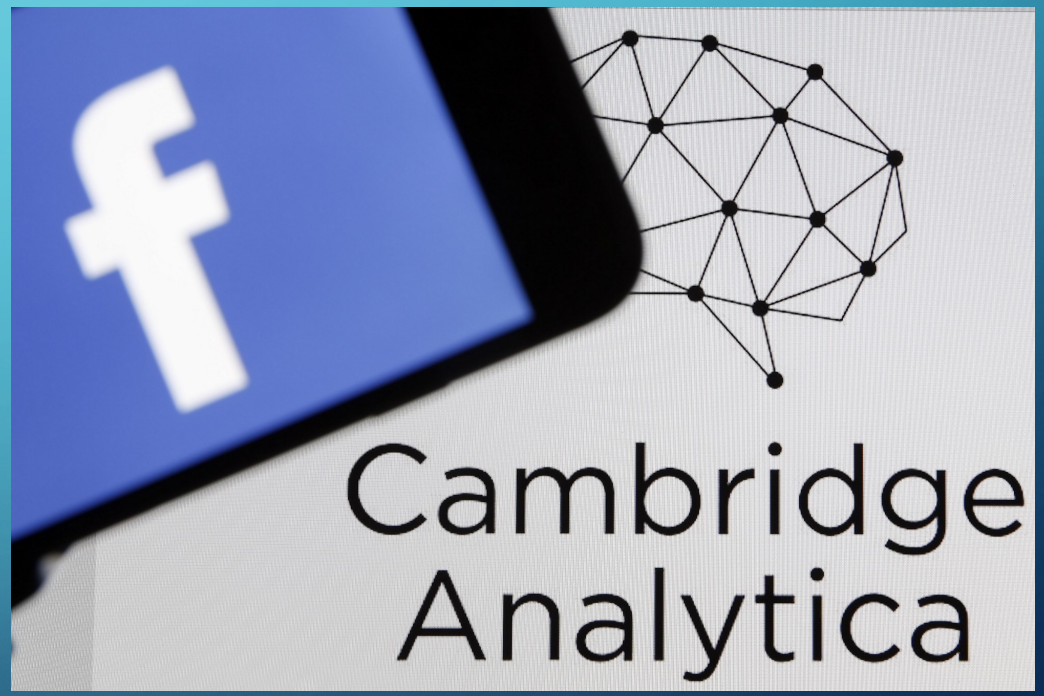
HOMO DEUS
YUVAL NOAH HARARI

«Cosa accadrà alla società, alla politica e alla vita quotidiana quando algoritmi non-coscienti, eppure altamente intelligenti, ci conosceranno meglio di quanto noi conosciamo noi stessi?»

OPINIONI?

2. MANIPOLAZIONE DEL COMPORTAMENTO

- Poiché software di analisi dati possono ricreare profili più o meno dettagliati di utenti (gruppi di utenti), questi software possono anche fornire informazioni utilizzabili per indirizzare scelte più o meno consapevoli.



INDIRIZZARE SCELTE INCONSAPEVOLI

BOLLE INFORMATIVE

- Una **bolla informativa** è un personale ecosistema di informazioni che viene soddisfatto da alcuni algoritmi.
- Per fornirci prodotti apprezzabili, diversi algoritmi inizieranno a selezionare e mostrarci oggetti che statisticamente sono di nostro gradimento.
- Questo potrebbe causare un isolamento informativo, laddove le informazioni che troveremo saranno sempre in linea con i nostri interessi e dunque limitanti per lo sviluppo di una visione plurale delle cose.

COSTRUZIONE SOFISTICATA DI INGANNI

https://thispersondoesnotexist.com/

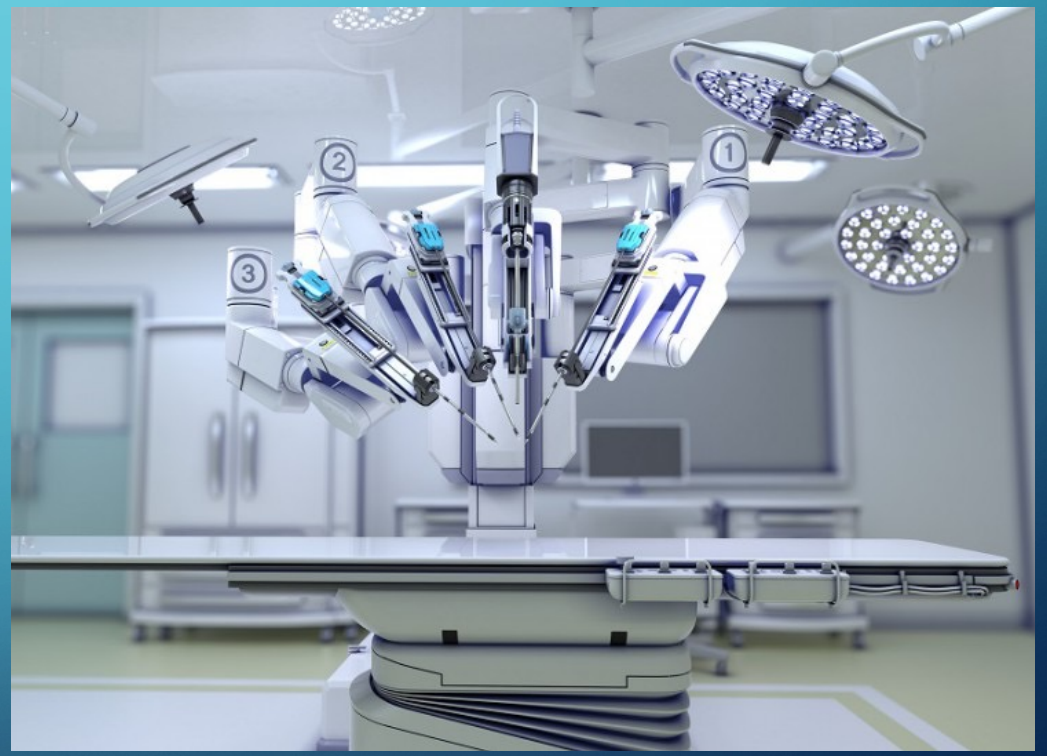
OPINIONI?

OPACITÀ DEI SISTEMI IA

- Le tecnologie informatiche sono sempre più coinvolte nei processi decisionali di attività umane.
- Quando i dati sono abbondanti, ma è difficile giungere ad un modello preciso che spieghi l'evolversi di una situazione, gli algoritmi di machine learning sono molto efficaci.

DECISIONI NON-IMPORTANTI

DECISIONI IMPORTANTI

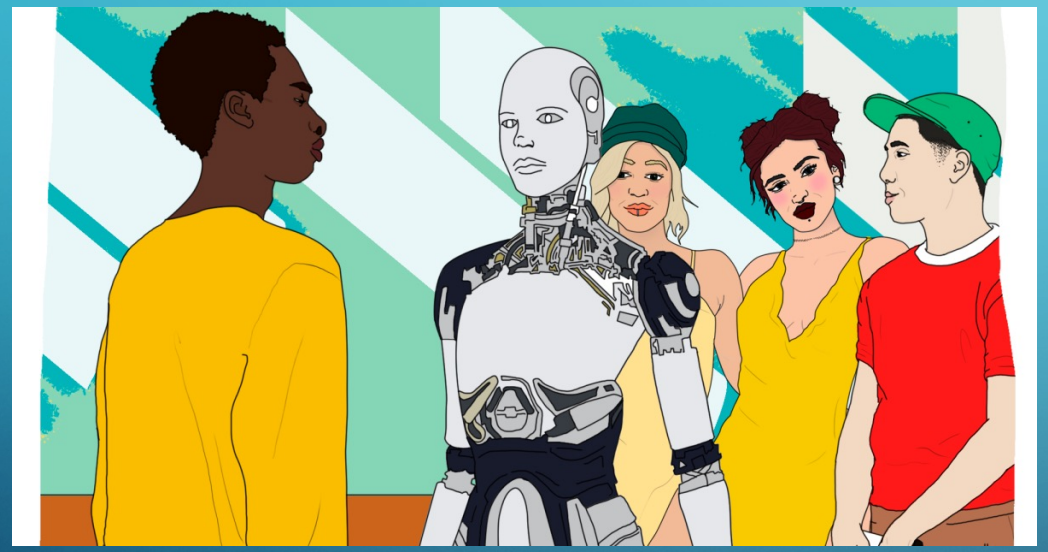


SEMPLICITÀ O PRECISIONE?

- Gli algoritmi di machine learning che costruiscono modelli a partire dai dati bruti utilizzano numerose variabili e le legano tra loro utilizzando complesse formule. Il risultato, spesso, sono modelli non lineari del fenomeno, con una buona precisione di risultato, ma quasi impossibili da comprendere per un essere umano (compresi gli ingegneri che hanno costruito l'algoritmo!!!).
- Dall'altro lato, abbiamo algoritmi di machine learning basati sulle white-box che forniscono risultati semplici da interpretare, ma, sovente, con un potere predittivo inferiore.

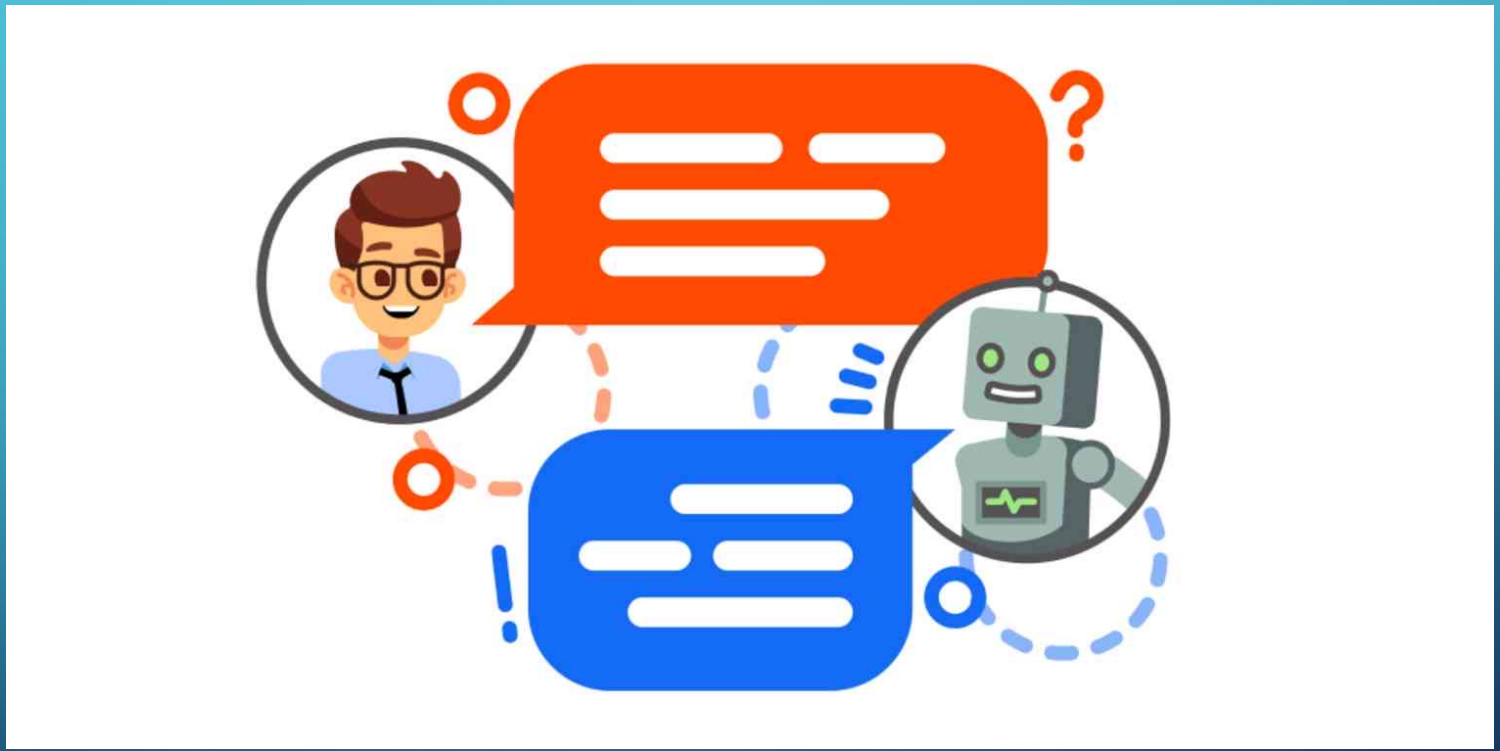
OPINIONI?

BIAS DELLE IA



- È facile cadere vittime della falsa credenza che gli oggetti tecnologici possano essere neutrali su diverse questioni importanti come il razzismo, il sessismo e i bias culturali.
- Tuttavia, questo non è sempre il caso.

TAY, CHATBOT MICROSOFT

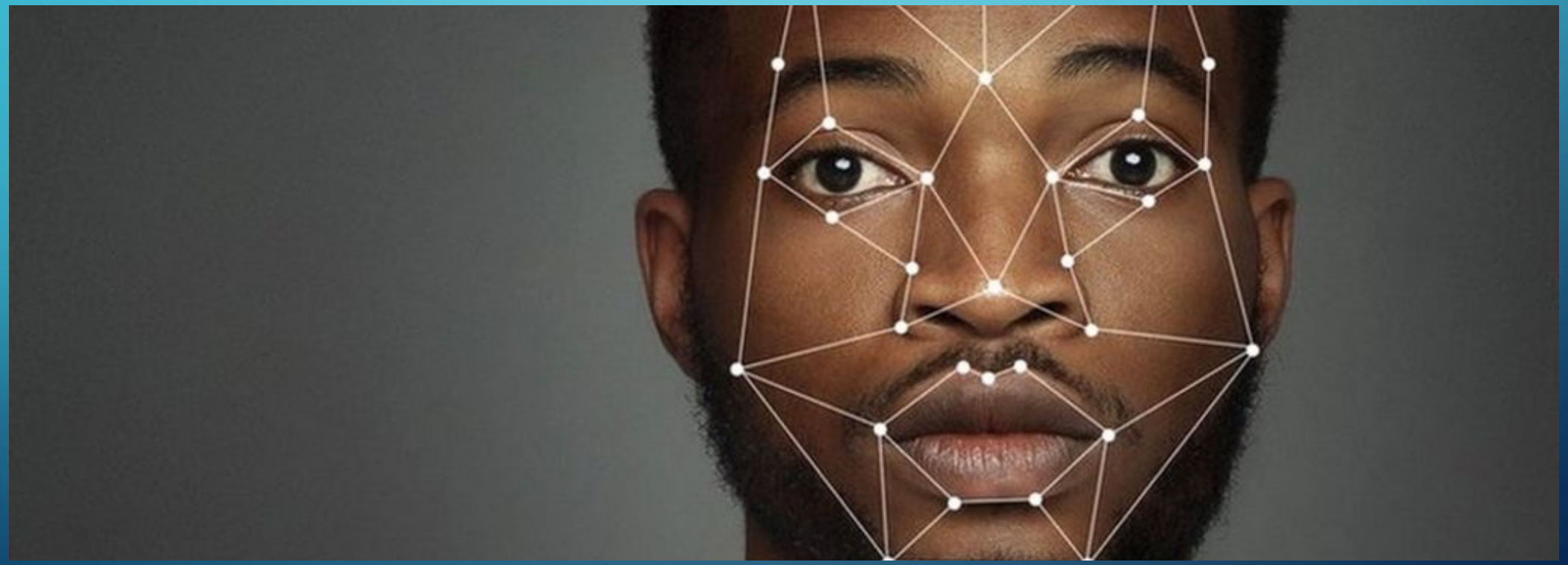


SCELTA ED INTERPRETAZIONE DEI DATI

- I bias che accompagnano gli algoritmi di machine learning sono spesso legati all'insieme di dati che viene scelto per addestrate uno specifico algoritmo. Se l'insieme dei dati non viene scelto accuratamente e se non si attua un importante lavoro di interpretazione dei risultati, c'è il rischio che le decisioni prese dal sistema siano soggette a dei bias non riconosciuti.

CORRELAZIONI SPURIE

BIAS GIÀ PRESENTI NEI DATI



OPINIONI?

ETICA DELLE MACCHINE

ETICA DELLE MACCHINE: COS'È?

- L'etica delle macchine è l'etica applicata dalle macchine come soggetti e non più come oggetti utilizzati dall'uomo.
- Il problema principe dell'etica delle macchine è quello di riuscire a garantire che le macchine si comportino in modo etico quando esse si interfacciano con altri agenti (siano essi esseri umani o altre macchine).
- Esistono vari modi di raggiungere questi obiettivi e diverse tematiche sono centrali nel comprendere bene cosa è necessario per avere un'etica delle macchine.

DARE LEGGI AD UNA MACCHINA

- Tra le possibilità a nostra disposizione per concettualizzare l'idea delle macchine etiche, c'è l'idea di inserire alcune leggi comportamentali all'interno del codice di una macchina.
- Seguendo tali leggi, la macchine dovrebbe comportarsi in modo etico.

COMPORTAMENTI INASPETTATI: IL COLLEZIONISTA DI FRANCOBOLLI

AGENTI MORALI ARTIFICIALI

- Un agente morale artificiale è un qualunque agente artificiale al quale vengono attribuiti diritti e responsabilità. Per poter parlare di macchine etiche, è necessario poter parlare di agenti morali artificiali.
- Questo poiché per poter attribuire un'etica ad una macchina, è necessario poterle attribuire delle responsabilità e dei diritti (tra i quali quello di potersi comportare diversamente, cioè di essere autonoma).

MACCHINE RESPONSABILI

- Alla base del concetto di responsabilità vi è la possibilità di attribuire il valore di una certa azione ad un certo agente.
- Per avere una macchina etica, è necessario che a tale macchina possa essere attribuita direttamente una responsabilità, piuttosto che distribuire tale responsabilità tra gli agenti che hanno contribuito alla sua costruzione.

ETICA E LEGGE A BRACCETTO

- Al centro del discorso sulle macchine etiche vi è dunque anche la giurisprudenza, poiché sono necessarie leggi che vincolino la responsabilità delle azioni di un determinato agente artificiale.

OPINIONI?

DIRITTI E DOVERI DELLE MACCHINE

- Se la responsabilità delle macchine è centrale, allora lo divengono anche i diritti e i doveri di tale macchina.
- Non solo: la libertà di agire di tale macchina è fondamentale.
- Se una macchina può agire in un solo modo, possiamo parlare di responsabilità della macchina nel caso in cui essa si comporti in maniera non consona?

OPINIONI?

SINGOLARITÀ

TEMI E PROBLEMI DI UNA INTELLIGENZA ARTIFICIALE GENERALE.

COS'È LA SINGOLARITÀ

- Uno degli obiettivi degli studi sull'IA è quello di creare una intelligenza artificiale generale (AGI: Artificial General Intelligence).
- La singolarità indica quell'istante temporale in cui l'uomo creerà un'intelligenza artificiale che possieda un livello umano di intelligenza, rendendola quindi in grado di creare a sua volta altre intelligenze artificiali (questa volta più intelligenti dell'uomo) in un crescendo sempre più accentuato di *superintelligenze*.

DALLA STORIA DEGLI UOMINI ALLA STORIA DELLE IA

- Argomenti a favore dell'avvento della singolarità si basano sull'aumento esponenziale delle capacità di calcolo della macchine e sui passi avanti nella programmazione di tali macchine.
- Secondo questi autori, questa rapidità di sviluppo porterà ad un'eventuale perdita di controllo nella progettazione e nello sviluppo di IA.
- Tuttavia, non tutti gli autori concordano sulla reale possibilità dell'avvento di una singolarità.

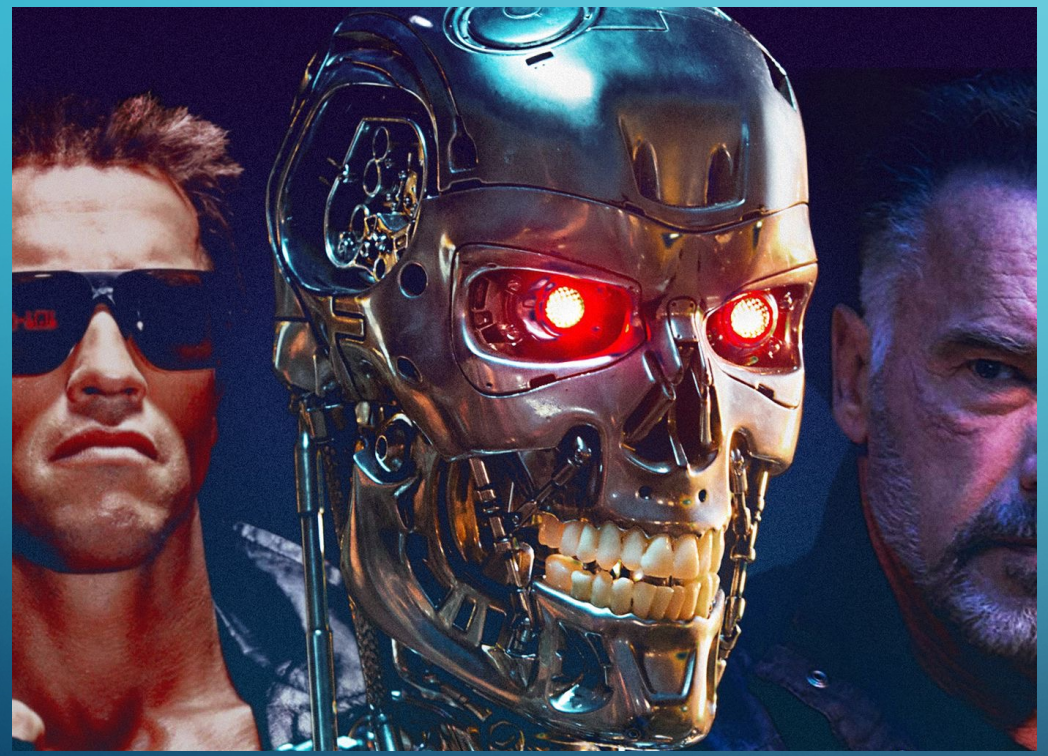
CRITICHE ALLA SINGOLARITÀ

- Esistono almeno tre livelli diversi di critica al concetto di singolarità:
  1. La singolarità potrebbe non essere concettualmente possibile;
  2. La singolarità potrebbe essere praticamente impossibile;
  3. La singolarità potrebbe essere impedita da eventi contingenti.

IPOTIZZARE NON SIGNIFICA CHE SIA POSSIBILE

https://www.youtube.com/watch?v=g0TaYhjpOfo

RISCHI DELL'AVVENTO DI UNA SINGOLARITÀ



- È comunque opportuno avere chiare le due problematiche principali legate all'avvento di una eventuale singolarità:

1. Rischi esistenziali per l'uomo;
2. Problema del controllo.

RISCHI ESISTENZIALI PER L'UOMO

- Il tema del rischio esistenziale per l'uomo all'interno del dibattito legato alla singolarità si concentra sulla possibilità o meno che l'avvento della singolarità sancisca la fine della vita umana. Tale possibilità si focalizza sul fatto che le super IA potrebbero avere obiettivi che confliggono con l'esistenza dell'uomo e, per questo motivo, che uno dei loro obiettivi strumentali sia quello di eliminare gli esseri umani.

RISCHIO ESISTENZIALE: OBIETTIVI ERRATI

- Il rischio esistenziale potrebbe emergere dall'attribuzione di un obiettivo errato alla superintelligenza.

RISCHIO ESISTENZIALE: COMPORTAMENTI IMPREVISTI

- Il rischio esistenziale potrebbe emergere da un comportamento inatteso atto a raggiungere l'obiettivo assegnato alla superintelligenza.

POSSIBILE SOLUZIONE: ALLINEARE GLI INTERESSI

- Tra le possibili soluzioni al problema del rischio esistenziale per l'uomo c'è la possibilità di allineare gli interessi dell'IA con gli interessi degli esseri umani.
- Assumendo che tra gli interessi degli esseri umani ci sia un interesse di preservazione della propria specie, sviluppando un adeguato allineamento tra gli interessi della superintelligenza e gli interessi degli esseri umani, tale superintelligenza non dovrebbe più risultare un rischio esistenziale per gli esseri umani.

OPINIONI?

PROBLEMA DEL CONTROLLO

- Il problema del controllo si allaccia alla questione dei comportamenti imprevisti di una superintelligenza.
- Si ha un problema di controllo quando una conseguenza inattesa di una determinata scelta non può più essere controllata dall'agente che ha fatto tale scelta.
- In tal senso, il problema etico del controllo delle superintelligenze è legato a come sia possibile mantenere il controllo di tali sistema anche una volta che essi hanno raggiunto livelli di intelligenza superiori a quelli degli esseri umani.

MATRIOSCHE DI INTELLIGENZE ARTIFICIALI SICURE

- Una delle idee principali per mantenere un controllo costante sulle intelligenze artificiali e su di una eventuale superintelligenza è quella di costruire intelligenze artificiali sicure, che, a loro volta, costruiranno altre intelligenze artificiali sicure.
- L'idea è dunque quella di controllare efficacemente i primi livelli delle IA, per poi avere un controllo anche sulle eventuali IA *post singolarità*.

OPINIONI?

ALTRI PROBLEMI INTERESSANTI: MACCHINA CHE NON VUOLE SPEGNERSI

- Esistono due tipologie di obiettivi: obiettivi principali e obiettivi strumentali.
- La preservazione della propria esistenza è uno degli obiettivi strumentali per eccellenza.
- Dunque, una qualunque superintelligenza che voglia raggiungere il proprio obiettivo principale, svilupperà l'obiettivo strumentale di non farsi spegnere.

HAL9000

https://www.youtube.com/watch?v=5rb42jHtDe4

ALTRI PROBLEMI INTERESSANTI: MACCHINA CHE NON VUOLE CAMBIARE

AUTONOMIA DELL'IA E DEI ROBOT

LE DOMANDE FONDAMENTALI ALLA BASE DELL'AUTONOMIA DELLA TECNOLOGIA

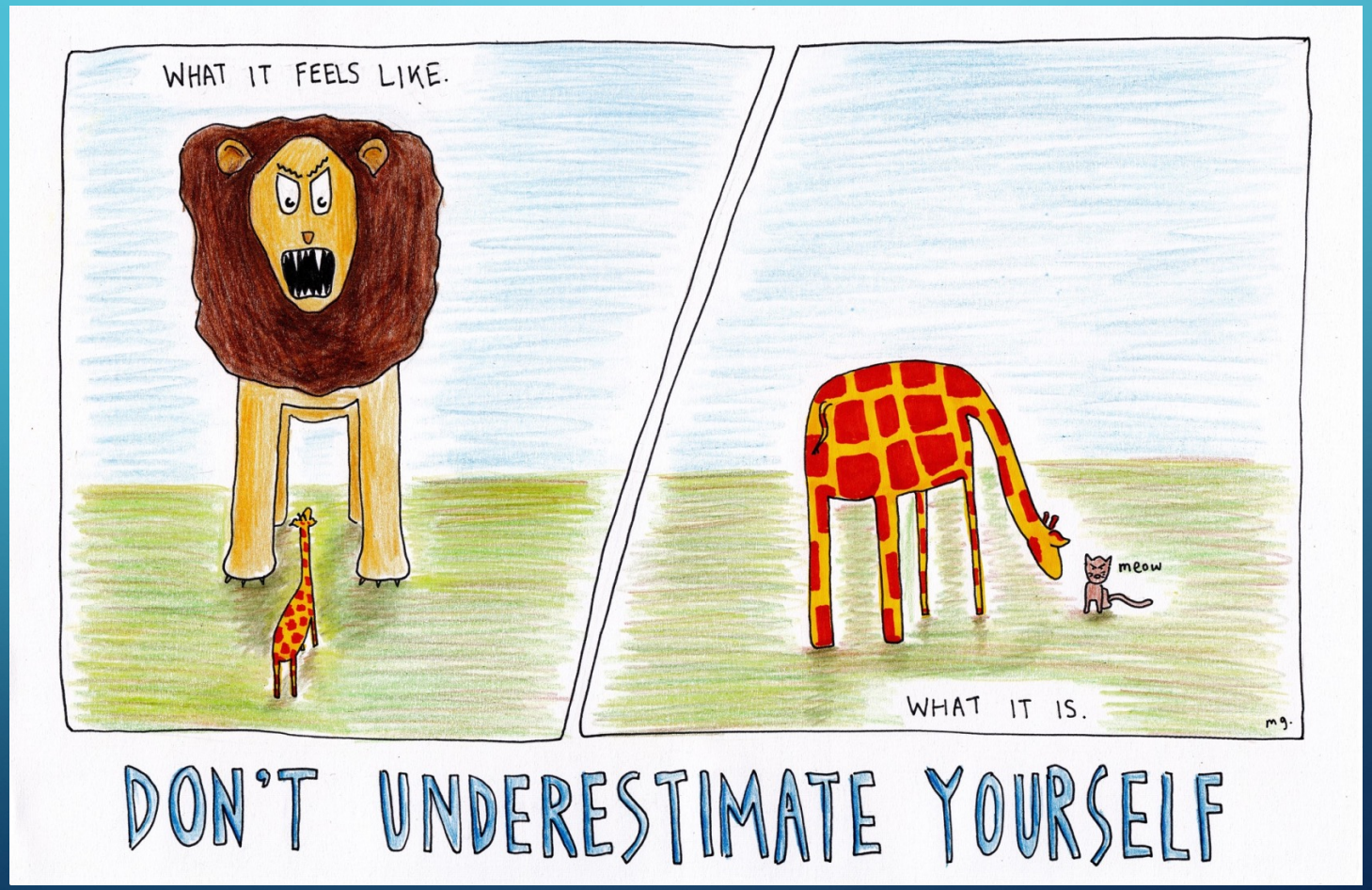
- Le questioni riguardanti le macchine etiche sono centrali nella comprensione del come e in che modo la tecnologia potrà essere autonoma. Se non saremo in grado di comprendere come un oggetto tecnologico si comporterà (o quali sono i limiti del suo comportamento), allora non avremo mai chiari quali sono i limiti dell'autonomia che possiamo concedere a tale oggetto tecnologico (sempre se sarà una nostra decisione porre tali limiti).
- Occorre quindi tener conto sia di elementi etici, che di elementi morali quando si parla di autonomia della tecnologia.

DOMANDA 1: HA SENSO PREOCCUPARSI?

- La capacità umana di creare oggetti tecnologici realmente autonomi e in grado di essere considerati come macchine etiche è ancora limitata: vale dunque la pena di preoccuparsi di questioni etiche e morali rispetto a tali oggetti tecnologici?

RISPOSTA A: PREVENIRE È MEGLIO CHE CURARE

RISPOSTA B: NON SOTTOVALUTIAMO LE CAPACITÀ INGEGNERISTICHE UMANE



DOMANDA 2: AUTONOMIA E REGOLE ETICHE POSSONO COESISTERE?

- Se l'obiettivo è implementare regole etiche all'interno di un programma di un oggetto tecnologico, in maniera tale che la sua autonomia non sia un pericolo per gli esseri umani, possiamo davvero parlare di autonomia di tale oggetto tecnologico?

ANTITESI:
ETICA E AUTONOMIA DELLA TECNOLOGIA

DOMANDA 3: COS'È UN'ARMA AUTONOMA?

- Poiché appare evidente che implementare un'etica all'interno dei codici degli oggetti tecnologici limiti grandemente la loro capacità di mostrare comportamenti autonomi, ma, allo stesso tempo, è stato sottolineato che gli esseri umani sono già in grado di creare oggetti tecnologici autonomi, ha senso chiedersi cosa siano realmente questi oggetti tecnologici autonomi di cui stiamo parlando.
- Abbiamo già discusso su una classe particolare di oggetti (le auto autonome). Oggi discuteremo di un'altra classe (le armi autonome).

ARMI AUTONOME

- Un’arma autonoma è definita come un’arma in grado di infliggere danno ad un avversario senza che sia necessario un intervento diretto di un essere umano.
- Esistono dunque diverse tipologie di armi autonome e non tutte sono particolarmente sofisticate.

POSSIAMO SPINGERCI OLTRE?

- I robot militari sono una sottoclasse delle armi autonome.
- Oltre ad essere armi autonome, essi sono anche in grado di percepire e manipolare il proprio ambiente, possedendo in tal senso un minimo grado di autonomia nelle proprie scelte.

ESISTONO DUNQUE DUE GRADI DI AUTONOMIA

- Il primo grado di autonomia è legato al fatto che alcune armi sono in grado di 'valutare' se è giunto il momento di agire e, data tale valutazione, procedono in autonomia ad arrecare un danno all'avversario.
- Il secondo grado di autonomia è legato al fatto che tali armi siano in grado di compiere tali valutazione in maniera attiva e non più in modo passivo, attendendo che l'avversario compia determinate azioni (e.g., cadere in una trappola).

DOMANDE O SUGGERIMENTI?

È DAVVERO COSÌ RILEVANTE?

- Il problema è di estrema rilevanza. Non tanto per il fatto che tali oggetti militari autonomi sono in grado di agire in maniera attiva nel proprio ambiente, ma, soprattutto, perché la loro presenza in ambito militare è gradualmente sempre più rilevante.
- In scenari dove le decisioni sono basati su enormi quantità di dati e sono richieste capacità di analisi di tali dati in tempi brevi, gli oggetti militari autonomi saranno sempre più chiamati a fare valutazioni e a proporre piano di azione.

FARE ALTRIMENTI?

- Vero è che le decisioni ultime sono spesso lasciate al giudizio di un essere umano. Possiamo però davvero parlare di libertà di scelta di tale essere umano?
- In questi scenari, gli esseri umani sono parzialmente deresponsabilizzati, poiché sono gli oggetti tecnologici a fornire le raccomandazioni di scelta. Allo stesso modo, la libertà dell'essere umano di scegliere altrimenti è messa seriamente in discussione, poiché un'errata valutazione verrebbe contrapposta alla raccomandazione di tali oggetti tecnologici.

ETICA MILITARE DEI ROBOT AUTONOMI

- Accettata l'inevitabilità del graduale affidamento delle scelte militari agli oggetti tecnologici autonomi, è possibile determinare in quali circostanze è consentito uccidere per un robot?
- Inoltre, è possibile limitarne le libertà di azioni violente ed omicide in tutte le altre situazioni non previste?

UTILIZZARE LE REGOLE GIÀ ESISTENTI

- Nel progetto di Arkin, l'uso delle regole attualmente utilizzate in campo militare può fornire una buona base per costruire un sistema di regole per gli oggetti tecnologici autonomi.
- Anzi, dal punto di vista di Arkin, gli oggetti tecnologici autonomi sarebbero in grado di seguire tali regole in maniera più efficace e precisa rispetto agli esseri umani.

TECNOLOGIA MIGLIORE RISPETTO ALL'UOMO?

- Arkin difende la superiorità degli oggetti tecnologici rispetto agli essere umani per due ragioni:

1. Tali oggetti tecnologici non provano emozioni che possono determinare comportamenti istintivi errati.
2. Tali oggetti tecnologici rispetteranno sempre le regole e seguiranno sempre gli ordini (a meno che questi ultimi non siano in contrasto con le regole generiche di comportamento).

DOMANDE O SUGGERIMENTI?

MORALE E STRATEGIA DISCIPLINANTE

- Il progetto di Arkin ha un respiro molto più ampio di quello che si possa pensare.
- Da un lato, tale progetto vuole condizionare la capacità di scelta degli oggetti tecnologici, senza costruire una vera e propria morale, ma imponendo una serie di vincoli ai possibili comportamenti di tali oggetti.
- Dall'altro lato, poiché tali regole limiteranno i comportamenti dei 'soldati', limiteranno anche le capacità di comportamento dei controllori, spingendoli a dover necessariamente seguire regole prescritte e, si spera, moralmente giustificate.

AGENTI NON TANTO AUTONOMI?

- Pensare all'etica di tali oggetti tecnologici seguendo le idee di Arkin pone però un problema. Seguire meccanicamente regole imposte dall'alto, senza aver modo di scegliere se seguirle o meno, pone seri dubbi sul possesso di autonomia di tali oggetti tecnologici.
- Dunque la mancanza di oggetti tecnologici realmente autonomi potrebbe non essere solamente un problema tecnico, ma un limite progettuale coscientemente scelto dagli esseri umani che questi oggetti li costruiscono.

DOMANDE O SUGGERIMENTI?

POSSIAMO AGGIUNGERE ALTRI DATI?

- Siccome le scelte di questi oggetti tecnologici sembrano essere vincolate al sistema di regole implementate negli algoritmi di tali macchine, è possibile aggiungere ulteriori elementi valutativi all'interno della banca dati di tali oggetti in modo tale che tali scelte siano più consapevoli?
- Potrebbe, ad esempio, un robot valutare se una determinata azione possa influenzare il comportamento futuro di un individuo e dunque scegliere se lasciarlo vivere, piuttosto che ucciderlo?

REGOLE CHE CAMBIANO

- Un ulteriore problema è quello del cambiamento delle regole in alcuni contesti o a seguito di alcuni avvenimenti.
- Occorre dunque avere la possibilità di modificare le regole seguite dagli oggetti tecnologici o permetter loro, almeno, di agire senza seguire tali regole.
- Fin dove possiamo però spingerci nel permettere questo discostamento dalle regole? Quanto deve poi esser rigida la procedura che lo permetta?

DOMANDE O SUGGERIMENTI?

CONCLUSIONI

GRAZIE PER L'ATTENZIONE

Swideas

SIMORA
RAZVOJNA AGENCIJA
SISAČKO MOSLAVAČKE ŽUPANIJE

Lai-momo

1506
UNIVERSITÀ
DEGLI STUDI
DI URBINO
CARLO BO

eurada

OBEC

One Block for Educational Credentials (OBEC)
2020-1-SE01-KA204-077803

Co-funded by the
Erasmus+ Programme
of the European Union