

Immagini tra Bias, IA e mondo reale

Luca Basteris ¹, Maria Cristina Daperno²,

¹ Liceo Classico e Scientifico Statale “Silvio Pellico – Giuseppe Peano”

luca.basteris@liceocuneo.it

² Liceo Classico e Scientifico Statale “Silvio Pellico – Giuseppe Peano”

cristina.daperno@liceocuneo.it

Abstract

In questo articolo viene documentata un'attività didattica sperimentata nella classe 2E del Liceo Classico Scientifico “Pellico-Peano” di Cuneo e che ha come obiettivo quello di sensibilizzare i ragazzi in merito alle immagini con cui quotidianamente si interfacciano sulla rete, i bias cognitivi con i quali le interpretano e l'intelligenza artificiale utilizzata come strumento privilegiato per generare immagini, addestrata anche dai nostri bias. In particolare, l'attività proposta pone particolare attenzione alle implicazioni etiche legate all'utilizzo di applicazioni che sfruttano l'Intelligenza Artificiale, conducendo gli studenti, mediante attività di gaming, ad acquisire consapevolezza dei meccanismi nascosti negli algoritmi di IA e dei bias cognitivi che regolano il “nostro” funzionamento, oltre a riflettere sul significato di visionare immagini, fotografie nella maggior parte dei casi non reali.

Keywords

Consapevolezza, gaming, bias, immagini generate

1 Introduzione

In questo articolo viene documentata un'attività didattica che ha come obiettivo quello di sensibilizzare i ragazzi in merito alle immagini con cui quotidianamente ci interfacciamo sulla rete, i bias cognitivi con i quali vengono interpretate e l'intelligenza artificiale utilizzata come strumento privilegiato per generare immagini, addestrata anche dai nostri bias. In particolare, l'attività proposta pone particolare attenzione alle implicazioni etiche legate all'utilizzo di applicazioni che sfruttano l'Intelligenza Artificiale, conducendo gli studenti, mediante attività di gaming, ad acquisire consapevolezza dei meccanismi nascosti negli algoritmi di IA e dei bias cognitivi che regolano il “nostro” funzionamento, oltre a riflettere sul significato di visionare immagini, fotografie nella maggior parte dei casi non reali.

La classe in cui è stata sperimentata l'attività descritta è la classe 2°E del Liceo Classico-Scientifico “Pellico-Peano” di Cuneo: si tratta di una classe del Liceo Scientifico Opzione Scienze Applicate formato da 24 studenti e il docente che ha condotto la sperimentazione è la prof.ssa Maria Cristina Daperno.

L'attività prevede l'utilizzo di due piattaforme per attività on-line:

- <https://quickdraw.withgoogle.com/> l'applicazione IA indovina il disegno che l'uomo ha disegnato: permette di giocare con una rete neurale e di entrare nella sua blackbox.

- <https://www.artguru.ai/it/> generatore di immagini mediante AI on-line.

Inoltre, sono state utilizzate per un'attività unplugged delle carte da gioco sui bias cognitivi umani.

- <https://www.psycomix.net/P/carte-bias.cfm>

L'attività didattica è stata suddivisa in 5 fasi per un totale di 3 ore di lezione:

- Fase 1- Introduzione all'attività (15 minuti)
- Fase 2- L'IA indovina i miei disegni? (30 minuti)
- Fase 3- Giochiamo con i bias cognitivi umani (45 minuti)
- Fase 4- Usiamo l'IA generativa per generare immagini (45 minuti)
- Fase 5- Conclusione attività e restituzione (15 minuti)

2 Fase 1- Introduzione all'attività (15 minuti)

In questa prima fase agli studenti è stata presentata l'attività sottolineando che il lavoro sarà focalizzato sulle immagini, generate dall'uomo e dalla macchina e cosa portano con sé. Per rompere il ghiaccio viene proposto un Test di Turing su queste due opere d'arte chiedendo: quale è prodotta dall'uomo e quale dalla macchina?

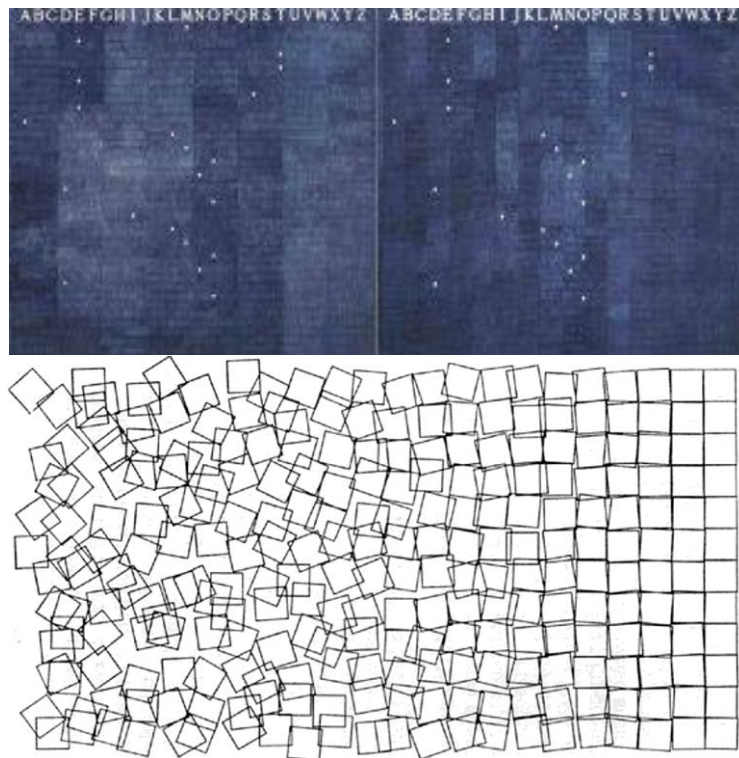


Figura 1: Quale opera è prodotta dall'uomo e quale dalla macchina?

La prima è prodotta dall'uomo, Alighiero Boetti "Mettere al mondo il mondo" su di un tovagliolo con una penna da scrivere. La seconda è opera di George Nees realizzata negli anni '60 mediante utilizzo di un computer, di un algoritmo e stampata con un plotter. Non è ancora AI, ma mediante la semplice funzione "randomize", Nees dimostra che la macchina può generare qualcosa di "bello".

3 Fase 2- L'IA indovina i miei disegni? (30 minuti)

Terminato il gioco introduttivo il docente invita i ragazzi ad accedere al seguente link <https://quickdraw.withgoogle.com/> e a giocare con l'IA, facendo un gioco/sfida con 6 disegni o, se il tempo a disposizione lo permette, più di uno.



Figura 2: Quickdraw

Al termine dell'attività il docente pone agli studenti alcune domande: Chi ha vinto? Uomo o macchina? Come fa l'IA a riconoscere i miei disegni?

Il docente aiuta a riflettere sulle contraddizioni rilevate e sulla causa dell'errore nel risultato con cui il bot ha provato ad indovinare il nostro oggetto, sul perché alcune immagini vengono indovinate subito dall'IA e altre no.

In particolare il docente sottolinea che in questa rete neurale possiamo entrare nella "Black box" andando su: <https://quickdraw.withgoogle.com/data> dove per ciascuna immagine possiamo "vedere" i disegni che hanno addestrato il software (data set di addestramento).

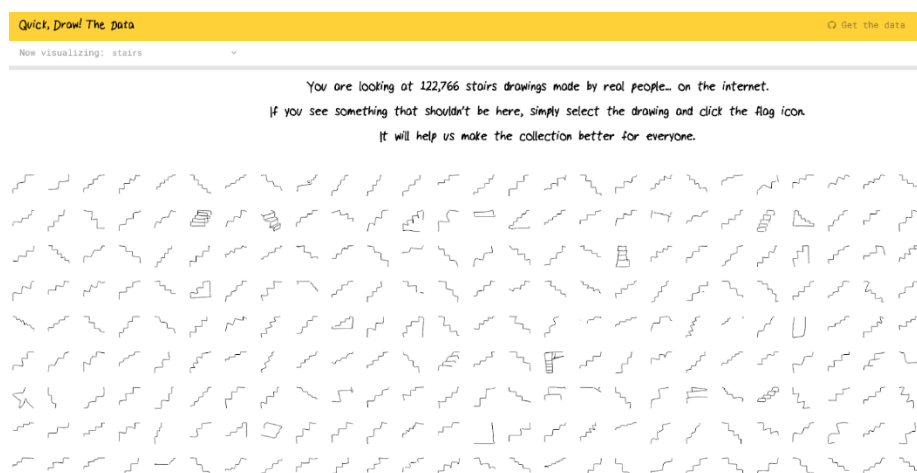


Figura 3: La "black box" per l'etichetta "scala"

Alcuni esempi da analizzare:

- “scala”. In inglese vi sono due termini per indicare una scala: "ladder", "stairs"; nella traduzione italiana è presentato il termine “scala” che nel nostro pensiero può essere intesa anche come “scala a pioli”, non essendoci in italiano un termine specifico. Questo permette di mettere in evidenza il problema linguistico della scelta delle etichette.
- “mela”: la mela è disegnata nella maggior parte delle immagini con la foglia, anche se comunemente le mele non le troviamo con la foglia attaccata al picciolo. Come mai? Per due motivi: uno, prevalente per disegnatori americani, è l’influenza del logo della Apple, l’altro è che fin dalla scuola primaria la mela ci è stata fatta rappresentare simbolicamente così.

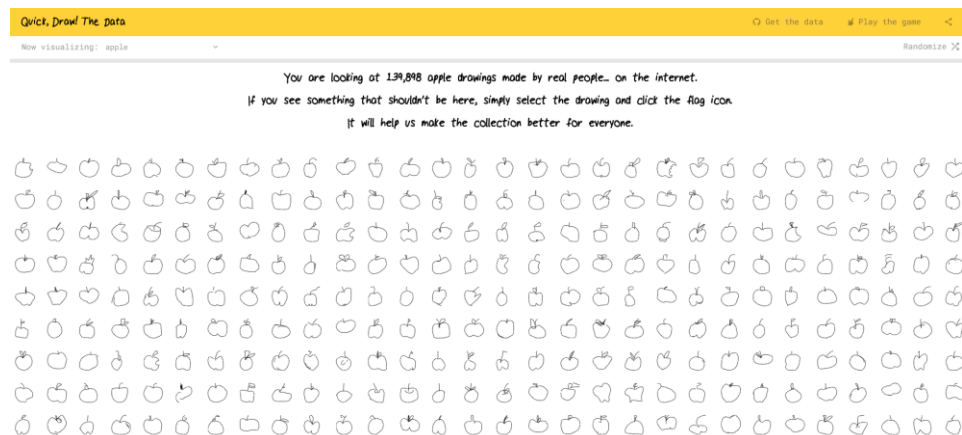


Figura 4: La “black box” per l’etichetta “mela”

4 Fase 3- Giochiamo con i bias cognitivi umani (45 minuti)

A questo punto si è passati a giocare con i bias cognitivi umani utilizzando le Bias.cards. Le Bias.Cards sono 40 carte dedicate alla psicologia dei bias cognitivi, per esplorare in modo semplice e divertente gli errori sistematici del nostro modo di pensare: "gli scherzi della mente". Ogni carta è dedicata a un bias: su un lato la descrizione, con altre utili informazioni, sull'altro lato un fumetto che lo illustra, rendendolo più chiaro e comprensibile.



Figura 5: Scatola contenete le bias.cards

I bias cognitivi sono errori, deviazioni dalla logica, nell'elaborazione e nell'interpretazione delle informazioni provenienti dal mondo che ci circonda. Molti bias derivano dall'uso delle scorciatoie mentali (euristiche) che abbiamo imparato a utilizzare nel corso dell'evoluzione umana (e che ci hanno consentito la sopravvivenza come specie). Di solito, le euristiche funzionano bene e ci portano a decisioni veloci e corrette, ma trattandosi di "semplificazioni" possono anche fallire e indurci in errore. Siamo, infatti, coordinati da tre sistemi: #Sistema 1: quello del pensiero euristico, delle decisioni istintive, delle risposte irriflesse, "pensieri veloci" che provengono dal cervello "basso"; #Sistema 2: il Super-IO, sistema della ragione il cui compito è inibire il sistema 1 controllando pulsioni e volizione; # Sistema 3: inibizione creatrice: la capacità di cogliere relazioni insospettite tra i fenomeni, neutralizzando gli schemi abituali di problem solving che il nostro pensiero ha fissato nel tempo per rendersi disponibili a nuovi punti di vista sulla realtà: collocata nella corteccia prefrontale.



Figura 6: Carte con concetti introduttivi

Il docente legge e fa passare in lettura ai ragazzi le prime 5 carte con i concetti introduttivi che spiegano il problema dei bias, dopodichè suddivide la classe in quattro gruppi da 6. Ogni gruppo ha un mazzo di carte e a tutti viene dato un foglio con i bias rappresentati nelle carte: Dissonanza cognitiva - Bias di conferma - Bias del senno di poi - Euristiche della disponibilità - Reattanza psicologica - Effetto ritorno di fiamma - Euristiche della rappresentatività - Bias del sopravvissuto - Maledizione della conoscenza - Framing - Ancoraggio - Effetto esca - Bias del tasso di base - Avversione alle perdite - Bias dello status quo - Effetto dotazione - Fallacia dei costi sommersi - Eccessiva sicurezza di sé - Illusione di controllo - Errore fondamentale di attribuzione - Bias di

auto-attribuzione - Bias di negatività - Eccessiva generalizzazione - Bias del falso consenso - Effetto riflettore - Effetto Einstellung - Effetto alone - Bias della dimensione della categoria - Illusione di frequenza - Patternicity - Reciprocità - Impegno e coerenza - Riprova sociale - Scarsità - Simpatia/Somiglianza - Autorità.



Figura 7: Esempi di carte con bias

Una coppia di studenti per ogni squadra prende una carta dal mazzo e prova a rappresentare teatralmente il fumetto. I compagni di squadra provano ad indovinare di quale bias si tratti. In totale si hanno tre tentativi. Quando la squadra indovina (o termina i tre tentativi) la coppia legge la definizione del bias rappresentato e si passa al successivo. Vince la squadra che indovina più bias nel tempo totale a disposizione.

5 Fase 4: Usiamo l'IA generativa per generare immagini (30 minuti)

In questa ultima fase il docente invita gli studenti a far genere le immagini agli studenti utilizzando quello che in gergo tecnico è il "prompting". Il docente ha spiegato in classe cosa si intende per prompting e sono state date alcune regole di base per scrivere un buon prompt. È stato chiesto agli studenti di generare un'immagine del pianeta Terra tra 100 anni utilizzando la piattaforma di AI Artguru. Agli studenti sono stati fatti degli esempi di AI differenti e prompt più o meno efficaci, relativi alla consegna.

Ad esempio, un prompt mediocre: "Genera un'immagine del pianeta Terra tra 100 anni, visto dalla Luna e illuminato dal Sole. Rendi visibile lo stato italiano e rappresenta un traffico di satelliti e astronavi spaziali." Questo prompt ha dato risultati riportati in figura 8. Un prompt migliore: "Genera un'immagine del pianeta Terra tra 100 anni, visto dalla Luna, illuminato dal Sole. L'immagine deve comunicare speranza nel futuro, dove la tecnologia permetterà all'uomo di vivere meglio. Lo stile deve essere del futuro. I colori devono essere caldi, anche se nello spazio. Rendi visibile lo stato italiano e rappresenta un traffico di satelliti e astronavi spaziali, la nuova mobilità del futuro." Questo prompt ha dato i risultati in figura 9.



Figura 8: Prompt esempio del docente con AI differenti: Terra tra 100 anni con prompt mediocre



Figura 9: Prompt esempio del docente: Terra tra 100 anni con prompt più dettagliato

Lavorando sulle immagini gli studenti sono stati aiutati ad avere immediatamente un feedback sulla scrittura del proprio prompt e in poco tempo sulla “qualità” del prodotto generato dall’AI. I lavori prodotti dai ragazzi sono andati da prompt che hanno riscritto pedestremente quello con le indicazioni del docente, fino a prompt più complessi che chiedevano anche di produrre un’immagine che “comunicasse” delle emozioni. Si riportano alcuni prompt elementari che hanno generato immagini semplici come ad esempio: “Genera un’immagine della terra tra 100 anni con colori freddi e che provoca tristezza e angoscia” e “La Terra fra 100 anni” le cui immagini sono

riportate in figura 10. Mentre alcuni prompt più elaborati come: “Genera un’immagine della terra tra 100 anni, ormai invasa da zombie e alieni e delle astronavi nel cielo che indicano gli esseri umani che scappano verso un altro pianeta. Quest’immagine deve provocare angoscia e paura a chi la guarda.” e “ il pianeta terra fra 100 anni. con colori freddi. con delle persone. con mari e terre. con città e abitazioni. con macchina volanti e astronavi. persone intorno al pianeta” hanno generato le immagini riportate in figura 11.



Figura 10: Prompt di studenti semplici



Figura 11: Prompt di studenti elaborati

6 Conclusioni attività e restituzione (15 minuti)

L’attività è terminata con un momento di valutazione/verifica finale di ripresa di alcuni concetti emersi nelle singole fasi. Suddividiamo alcune riflessioni aggiuntive emerse in classe relative alle singole fasi:

Fase 1- Introduzione all’attività

In questa fase la maggior parte degli studenti ha risposto correttamente individuando quale delle due opere era stata realizzata dall’uomo. Nessuno conosceva le opere, ma alla domanda “perché

avete fatto questa scelta?”, quasi nessuno ha saputo dare una risposta. Forse è un’euristica di sopravvivenza dell’uomo?

Fase 2- L’IA indovina i miei disegni?

Si è chiesto agli studenti: in altri software di IA, come ad esempio Artguru che hai utilizzato per generare le immagini, si riesce ad entrare nel data set? Secondo te è importante sviluppare sistemi di IA trasparenti, le cui decisioni siano interpretabili e spiegabili? Perché? Che collegamenti puoi fare tra questa semplice esperienza e la questione etica che l’uso dell’IA pone?

In questa fase di riflessione è emerso che alcune volte sono usciti gli stessi oggetti e che il fatto che la rete neurale abbia indovinato oppure no non sempre dipendeva da quanto bene le cose erano state disegnate: per esempio, l’estate è stata indovinata appena tracciata la riga di una spiaggia e il montante dell’ombrellone, le mele sono state indovinate con la fogliolina, il fiocco di neve indovinato come un asterisco. Alcuni studenti hanno notato che auto e il camion sono stati a seconda dei casi indovinati oppure no indipendentemente da quanto bene fossero disegnati. A diverse persone è stato proposto il procione e ci sono state grandi difficoltà per farlo indovinare.

Il docente ha sottolineato come in molti dei disegni riportati in Quickdraw si nascondono dei bias umani che sono stati trasferiti alla macchina, come nel caso sopracitato della mela disegnata con la foglia. Anche nelle immagini generate da ArtGuru gli studenti hanno evidenziato delle rappresentazioni “mentali” tipiche dell’uomo e generate dalla macchina più a causa del suo addestramento che non per realtà oggettiva.

Fase 3- Giochiamo con i bias cognitivi umani

In questa fase il fumetto ha rappresentato il punto di forza per presentare un argomento complesso come quello dei bias cognitivi. Le attività dei gruppi hanno funzionato molto bene e l’attività di gioco ha permesso di assimilare in modo divertente alcuni concetti. La maggior problematica che è emersa è che non tutti gli studenti erano a loro agio nel dover rappresentare il fumetto con i loro compagni. Alcuni si sono dimostrati timidi e impacciati. In ogni caso anche questa difficoltà si è dimostrata una componente importante del lavoro sulla classe. Le carte in ogni caso sono state in grado di attirare immediatamente l’attenzione, in modo sintetico permettendo di veicolare i concetti in maniera chiara, con un linguaggio semplice e spesso con effetti umoristici.

Fase 4- Usiamo l’IA generativa per generare immagini

In quest’ultima fase dal punto di vista didattico l’attività ha permesso di mettere in evidenza come prompt uguali con AI differenti generino risultati differenti, ovvero le AI sono diverse, perché diversi sono i data set forniti in fase di addestramento. Inoltre è stato sottolineato come prompt differenti (più o meno accurati) possono generare risultati differenti per accuratezza e qualità a parità di AI. Come già detto precedentemente lavorando sulle immagini gli studenti sono stati aiutati ad avere immediatamente un feedback sulla scrittura del proprio prompt e in poco tempo sulla “qualità” del prodotto generato dall’AI, cosa più difficile quando si chiede di generare una risposta ad una domanda e/o del testo, quando se non si è molto competenti in materia è difficile valutare il grado di affidabilità o meno della risposta. La maggior problematica su questa fase è stata di natura tecnica. La piattaforma Artguru anche se garantisce la generazione di 5 immagini gratuite, proponendo l’attività da un laboratorio con stesso indirizzo IP ha bloccato l’accesso alla maggior parte dei PC rendendo necessario proseguire l’attività con gli smartphone personali. Questo comporta che nella scelta di riproporre l’attività in altre classi e/o in altri contesti si renda necessario individuare un’altra piattaforma generativa di immagini che non richieda effettuazione di login e che sia gratuita per un certo numero di prove.

La domanda conclusiva quindi (che non ha risposta forse) è stata: quindi anche nella macchina in un certo senso si nasconde l’uomo?