Hyeongjoo Kim, Dieter Schönecker (Eds.), *Kant and Artificial Intelligence*, De Gruyter, Berlin-Boston 2022, pp. 290, € 19.95, ISBN 9783111355696

Giulio Amore
Università degli Studi di Padova

La raccolta di saggi edita da Hyeongjoo Kim e Dieter Schönecker rappresenta una delle più interessanti pubblicazioni scientifiche tra quelle centrate sul rapporto tra la riflessione filosofica classica e le tematiche relative all'intelligenza artificiale. I saggi compresi all'interno del volume indagano in tre sezioni (filosofia teoretica, pratica ed estetica) i rapporti tra il pensiero kantiano e l'intelligenza artificiale, dal problema cognitivo a quello della coscienza, passando per dilemmi morali e capacità di provare sentimenti.

Il primo articolo, di Tobias Schlicht, intitolato *Minds, Brains, and Deep Learning: The Development of Cognitive Science Through the Lens of Kant's Approach to Cognition*, funge da capitolo introduttivo, andando a ripercorrere i tratti salienti delle differenti modalità attraverso le quali l'approccio kantiano al tema della conoscenza è stato influente e rilevante per lo sviluppo di paradigmi nell'ambito delle scienze cognitive, in particolar modo in funzionalismo, enattivismo e modellazione predittiva.

L'articolo cerca di sondare se sia possibile utilizzare Kant per comprendere in che modo avvenga la produzione di conoscenza in una mente ed inoltre se modelli ispirati allo schematismo kantiano possano essere applicati a sistemi artificiali, mostrando possibili sviluppi di ricerca futuri e limiti intrinseci a questi sistemi.

L'articolo seguente *The Apperception Engine* è a tutti gli effetti la descrizione dettagliata di una macchina kantiana. Il sistema realizzato, chiamato dall'autore Richard Evans *The Apperception Engine*, è il tentativo di riutilizzare "la psicologia a priori" kantiana come "architectural blueprint for a machine learning system" (p. 39). La scelta di utilizzare Kant è da Evans subito motivata: "Kant asks for the conditions that must be satisfied for the agent to have any possible cognition […] this is not an empirical psychological question about the processes that

human beings happen to use, but rather a question of *a priori* psychology: what must a system – any physically realised system at all – do in order to achieve experience?" (p. 44).

Evans, nella prima parte del testo, descrive le condizioni da soddisfare per raggiungere l'obiettivo fondamentale al fine di produrre conoscenza, ovvero unificare l'esperienza. Nel farlo, l'autore cerca di estrapolare dalla psicologia kantiana una forma molto asciutta di queste condizioni, così da poterle formalizzare al fine di implementarle all'interno di un sistema informatico.

La seconda parte dell'articolo invece espone l'applicazione vera e propria della struttura descritta all'interno di un computer ed è costituita in larga parte da esperimenti e dal commento degli stessi, la cui piena comprensione richiede tuttavia la conoscenza di alcuni elementi chiave del *machine learning*.

Il terzo articolo, scritto da Sorin Baiasu, intitolato *The Challenge of (Self-)Consciousness: Kant, Artificial Intelligence and Sense-Making*, ruota attorno al cosiddetto problema della rappresentazione in ambito cognitivo. Dopo aver ripercorso brevemente alcune posizioni sul problema, il paper si sposta sui tentativi di risolverlo – soprattutto in ambito informatico – grazie all'analogia con la distinzione kantiana tra sensibilità e intelletto.

Sulla base di questa distinzione, molti programmi sono riusciti ad ottenere livelli considerevoli di successo in termini di ragionamento proposizionale, in particolare sfruttando la possibilità di ridurre i dati "sensibili" (*low-level perception*) a stringhe di dati, analizzabili dalle facoltà "intellettive" (*high-level perception*) di queste macchine.

Tuttavia, l'autore sottolinea che sebbene i risultati siano innegabili e molti di essi siano di assoluto valore, la pretesa a volte avanzata di aver di fronte dei veri e propri agenti cognitivi sia ancora tutta da provare, mancando di fatto un elemento fondamentale per un completo atto cognitivo (e dunque l'esistenza di un agente cognitivo): l'io penso.

Il quarto ed ultimo saggio relativo alla filosofia teoretica riguarda da vicino il problema della definizione in termini filosofici di che cosa sia l'IA. Hyeongjoo Kim, in *Tracing the Origins of Artificial Intelligence: A Kantian Response to McCarthy's Call for Philosophical Help*, cerca di farlo a partire dalla distinzione tra idealismo trascendentale e realismo trascendentale,

evidenziando il discrimine fondamentale tra la concezione di intelligenza di McCarthy (realismo trascendentale) e quella di Kant (idealismo trascendentale). Sebbene entrambi gli autori, per Kim, riconoscano come tratto tipico della facoltà intellettiva una capacità combinatoria, ciò che li divide e traccia la differenza essenziale è la presenza di un'ulteriore facoltà, ovvero, la capacità di essere consci della propria stessa attività intellettiva.

Il quinto saggio, intitolato *A Kantian Perspective on Robot Ethics*, di Lisa Benossi e Sven Bernecker, apre la sezione dedicata alla filosofia pratica. Il fulcro del saggio ruota attorno ad una domanda: quali condizioni vanno soddisfatte affinché un robot possa esser considerato un agente morale? La questione è affrontata dagli autori attraverso due punti di vista: uno più neutrale e l'altro invece di ispirazione marcatamente kantiana. Se per una visione più ampia e generalmente detta "ottimista" parlare di robot come di agenti morali sembra essere quantomeno possibile – sebbene solo in presenza della cosiddetta *strong AI* – da un punto di vista kantiano è invece irrealistico. Questo perché, come gli autori fanno notare, per considerare un individuo un agente morale, è necessario che esso possa accedere alla legge morale e che sia autonomo nelle sue decisioni. Anche in relazione ad una supposta *strong AI*, questo scenario rimarrebbe comunque difficile a verificarsi, proprio in virtù dell'assenza di un'autonomia vera e propria. Il saggio si chiude con alcune considerazioni sulla desiderabilità di avere degli agenti morali robot, presentando con dovizia le diverse posizioni attuali sull'argomento.

Nel saggio seguente *Kant's Argument from Moral Feelings: Why Practical Reason Cannot be Artificial*, Dieter Schönecker analizza da un punto di vista kantiano la possibilità di pensare una ragion pratica artificiale. L'autore è sin da subito molto chiaro: dal momento che la ragion pratica kantiana si fonda sulla presenza di sentimenti morali, attraverso i quali l'essere umano valuta la validità della legge morale come imperativo categorico, e nessuna AI può avere sentimenti, essa non può avere nemmeno sentimenti morali. Questo, secondo Schönecker, è dovuto in larga parte all'assenza di un vero e proprio "io autocosciente" che possa effettivamente "sentire": "[a] feeling that is not felt is not a feeling; but for it to be felt there must be someone who feels it" (p. 184).

Nell'articolo numero sette, *Kant on Trolleys and Autonomous Driving*, Elke Elisabeth Schmidt dibatte intorno al famoso esperimento mentale del carrello ferroviario (Foot, 1978). L'autrice afferma che, a differenza della maggior parte delle interpretazioni, la decisione di Kant di fronte al dilemma sarebbe quella di non muovere la leva: non uccidere la singola persona sarebbe infatti un dovere morale superiore al dover salvare le altre cinque. Prevarrebbe dunque il carattere negativo di un divieto e non un principio di massimizzazione del bene. L'articolo è incluso all'interno della raccolta perché affronta un tipico problema che emerge nelle fasi di pianificazione e progettazione dei veicoli autonomi. Il saggio considera l'esperimento mentale di Foot utile per discutere parte delle problematiche dei veicoli autonomi, soprattutto se interpretato kantianamente. Sebbene i critici ne evidenzino la scarsa aderenza a scenari reali e la complessità di essi e degli attori in esso attivi, esso porterebbe comunque a risultati teoreticamente utili ai fini di risolvere problemi pratici.

Il dibattito intorno ai veicoli autonomi è ripreso anche nel saggio di Ava Thomas Wright (*Rightful Machines*). L'autrice parte dalla considerazione che le macchine altamente autonome debbano seguire una morale che non sia delle virtù, quanto più una morale della giustizia, basata sui doveri che tutti potrebbero accettare. Partendo dalla distinzione tra "*duties of right* ('legal' duties)" e "*duties of virtue* ('ethical' duties)" (p. 224), Wright argomenta che un'etica basata sul rispetto dei *duties of virtue* potrebbe essere troppo spesso soggetto di dispute all'interno dei vari casi particolari. Al contrario, una morale che seguisse come traccia fondamentale uno standard giuridico-legale rappresenterebbe un'opzione più consona ai veicoli autonomi, soprattutto se considerati all'interno di dilemmi morali come il già citato problema del carrello ferroviario (proposto anche da Wright). Inoltre, dispute derivanti da scenari reali analoghi a questi dilemmi si risolverebbero comunque in primo luogo all'interno di un contesto giuridico e di diritto pubblico.

La natura sociale del problema viene ripresa anche nel saggio seguente (*Partners, Not Parts. Enhanced Autonomy Through Artificial Intelligence? A Kantian Perspective*), nel quale Claus Dierksmeier sostiene la necessità di concentrarsi innanzitutto nella formulazione di un framework legale per le applicazioni basate sull'IA. Tuttavia, l'autore precisa che a partire da questi termini la moralità dei singoli software vada giudicata caso per

caso. Nelle conclusioni, va anche segnalato l'invito a realizzare progetti dal design partecipato, in modo tale da rendere maggiore l'impatto della società intera nello sviluppo delle varie piattaforme, garantendo una maggiore aderenza (ipotetica) ad una prospettiva kantianamente ispirata.

Nel decimo e ultimo saggio *On the Subjective, Beauty and Artificial Intelligence: A Kantian Approach*, che rappresenta l'unico articolo dedicato all'estetica, Larissa Berger riprende l'argomento relativo all'impossibilità per un'intelligenza artificiale di provare sentimenti. Questa carenza porta l'autrice a considerare l'esperienza estetica del bello, descritta a partire dalla riflessione kantiana sull'argomento, come impossibile: "[the] realm of beauty is foreclosed to AI" (p. 280).

La raccolta, presa nella sua interezza, risulta essere una delle più esaustive disamine delle questioni centrali all'interno del dibattito intorno all'intelligenza artificiale. L'approccio kantiano, seppur in alcuni passaggi non del tutto problematizzato, connette tuttavia coerentemente i vari contributi, ponendo in rilievo l'assoluto valore della riflessione filosofica classica anche all'interno di un contesto solo apparentemente estraneo ad essa.

Bibliografia

Philippa Foot, *The Problem of Abortion and the Doctrine of the Double Effect*, «Oxford Review», 5, 1967, pp. 5-15

Ulteriori recensioni dello stesso volume

H. Compston, *K. Hyeongjoo, D. Schönecker (Eds.), Kant and Artificial Intelligence, De Gruyter, Berlin-Boston 2022*, «Kantian Review», 2024, pp. 1-4