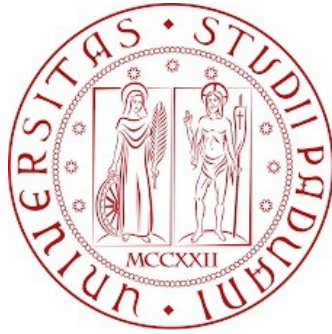




UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI FILOSOFIA, SOCIOLOGIA, PEDAGOGIA E PSICOLOGIA APPLICATA

CORSO DI LAUREA IN FILOSOFIA

“La deriva della tecnica: ottimismo del progetto e pessimismo della conseguenza in Turing e Anders”

Relatore:

Ch.mo Prof. Fabio Grigenti

Laureanda:

Erica Marangon

Matricola 2071656

- **Introduzione**

CAPITOLO 1 – L’OTTIMISMO DEL PROGETTO IN ALAN TURING.

- 1.1 – C’è qualcosa che una macchina non possa fare?
- 1.2 – “Macchine calcolatrici e intelligenza” di Alan Turing.
- 1.3 ---Turing e l’eresia del controllo delegato.
- 1.4 -- La dissoluzione del mistero: il funzionalismo e la riduzione logica dell’essere umano.
- 1.5- La smaterializzazione del soggetto: verso un’intelligenza senza corpo.

.

CAPITOLO 2 – GUNTHER ANDERS, LA TECNICA COME MINACCIA ESISTENZIALE.

- 2.1 -- La condizione umana nell’epoca tecnica: genesi e fenomenologia della “Vergogna Prometeica”
- 2.2 – Il dislivello prometeico: produrre senza comprendere.
- 2.3 – Il mondo come fantasma: la mediazione tecnica dell’esperienza.
- 2.4– La delega della responsabilità in “Noi figli di Eichmann”.
- 2.5 – La smaterializzazione dell’esperienza: il mondo come fantasma e l’uomo come pezzo di ricambio.
- 2.6 – L’inversione della libertà: lo scarto tra rappresentazione e produzione.
- 2.7– Il conformismo tecnico: l’uniformazione attraverso l’apparato.
- 2.8 – L’apocalisse senza apocalisse: la catastrofe come mutamento ontologico.

CAPITOLO 3 -- LA DERIVA DELLA TECNICA NEL PRESENTE: TURING E ANDERS E LA QUESTIONE.

- 3.1 -- Il malinteso come sintomo: bug tecnico o abisso ontologico?
- 3.2 – Dalla bomba atomica all’intelligenza artificiale: la logica dell’incontrollabilità.
- 3.3 – Il caso D-Syde: esperienza tecnica e coscienza del limite.

- **Conclusione**

Introduzione

Negli ultimi anni la domanda “Le macchine possono pensare?” è tornata molto attuale. Non riguarda solo l’informatica, ma anche la filosofia: quando affermiamo che qualcosa pensa, che cosa intendiamo davvero? E soprattutto: cosa cambia per noi, come uomini, quando sempre più attività mentali vengono affidate alle macchine?

Questa tesi inizia da qui e mette a confronto Alan Turing e Gunther Anders. Il titolo, “La deriva della tecnica: ottimismo del progetto e pessimismo della conseguenza in Turing e Anders”, mostra già l’idea centrale. Da una parte c’è la fiducia nella progettazione tecnica, dall’altra la preoccupazione per le conseguenze che la tecnica produce e produrrà nel tempo, spesso oltre le intenzioni iniziali.

Nel saggio “Computing Machinery and Intelligence” (1950), Turing evita di rispondere in modo astratto alla domanda “Le macchine possono pensare?” e propone invece un criterio pragmatico: l’“Imitation Game”¹. Sostanzialmente, sposta l’attenzione dall’essenza del pensiero al comportamento osservabile: se, in un dialogo, una macchina risponde in modo indistinguibile da un uomo, allora ha senso attribuirle una forma di intelligenza e pensiero. Questa posizione mostra un forte ottimismo del progetto, la convinzione che alcune capacità cognitive possano essere descritte, simulate e sviluppate in chiave tecnica. Già in “Intelligent Machinery” (1948), Alan Turing mostra come le macchine possano addirittura apprendere. Questo aspetto è essenziale perché fa capire che, per lui, l’intelligenza è un processo che può crescere e modificarsi.

Con Anders il discorso cambia radicalmente prospettiva. Nei due volumi “L’uomo è antiquato”, la tecnica non è colta come strumento utile, ma come forza storica che rischia di superare la nostra capacità di comprenderne e governarne gli effetti³. Il suo concetto di “dislivello prometeico” descrive proprio questo: siamo capaci di creare molto più di quanto riusciamo ad immaginare e a controllare sul piano etico e politico. In “Noi figli di Eichmann” Anders insiste sul concetto oramai problematico di responsabilità: in sistemi tecnici complessi, le persone possono limitarsi a “far funzionare” una parte del processo senza sentirsi responsabili delle conseguenze finali.

Per questo la tesi userà le due prospettive insieme:

1. Turing aiuta a comprendere come possiamo parlare di intelligenza delle macchine.
2. Anders aiuta a comprendere quali conseguenze tale realtà produce sull’umano e sulla responsabilità.

¹ A. M. Turing, “Computing Machinery and Intelligence”, *Mind*, 59(236), 1950, pp. 433–460.

² A. M. Turing, “Intelligent Machinery” (1948), poi in D. C. Ince (a cura di), *Mechanical Intelligence*, North-Holland, 1992.

³ G. Anders, “L’uomo è antiquato”, vol. I e II, Bollati Boringhieri.

L'ipotesi del lavoro è che la “deriva della tecnica” emerga proprio dallo scarto tra progetto e conseguenze; progettiamo strumenti per implementare capacità e controllo, ma gli effetti morali ed antropologici possono andare in una direzione completamente opposta come si sta dimostrando.

La struttura sarà questa: nel primo capitolo analizzerò la posizione di Turing, nel secondo quella di Anders. Nel terzo invece, metterò a confronto i due autori per mostrare punti di contatto, differenze e limiti reciproci. Il fine non è definire se una macchina pensa davvero, ma chiarire al meglio i livelli della domanda: livello tecnico-cognitivo e livello etico-antropologico.

Capitolo 1 – L'ottimismo del progetto in Alan Turing.

Capitolo 1.1 – C'è qualcosa che una macchina non possa fare?

Il punto di partenza della riflessione di Alan Turing non riguarda tanto l'ontologia della macchina in quanto tale, ma piuttosto l'individuazione di eventuali limitazioni invalicabili alle sue capacità operative. La domanda che attraversa tutto il suo pensiero è provocatoria e destinata a mettere in discussione molte certezze del senso comune: c'è qualcosa che una macchina non possa fare? Questa formulazione apparentemente semplice sfida l'idea, molto radicata non solo nel senso comune ma anche in parte della riflessione filosofica classica, che esistano domini dell'agire umano strutturalmente preclusi alla riproduzione meccanica. In questa visione, la tecnica non è colta come un semplice mezzo nelle mani dell'uomo, ma come un'entità capace di cancellare progressivamente il monopolio umano sulle facoltà cognitive. Se si immagina un sistema meccanico sufficientemente complesso, la distinzione tra ciò che è naturale e ciò che è artificiale diventa estremamente labile, quasi inconsistente dal punto di vista funzionale.⁴

Vale la pena sottolineare che storicamente il tentativo di replicare le capacità umane ha una lunga tradizione che anticipa di secoli la rivoluzione informatica del Novecento. Già nel pieno Settecento illuminista Jaques de Vaucanson, ingegnere e inventore francese, costruiva automi meccanici che imitavano non solo il movimento corporeo ma anche alcuni processi che al tempo sembravano richiedere capacità cognitive autentiche, come suonare uno strumento a fiato, o muovere le dita su una tastiera musicale in modo preciso seguendo uno spartito complesso. Questi dispositivi meccanici elaborati non erano affatto semplici giocattoli volti allo svago dalla nobiltà: erano veri e propri esperimenti filosofici, che ponevano in modo concreto la domanda metafisica su dove passa esattamente il confine tra imitazione convincente e realtà autentica, tra simulazione efficace e comprensione genuina. L'altra meccanica di Vaucanson, che sembrava mangiare, digerire, sollevava nel Settecento domande che anticipavano quelle di oggi; se un sistema riproduce tutte le funzioni osservabili di un organismo vivente, in che senso possiamo dire che non è davvero vivo?

Tuttavia, con il progresso tecnologico accelerato del Novecento questa differenza ontologica presunta è diventata sempre più problematica e difficile da sostenere filosoficamente. Azioni e comportamenti considerati naturalmente e specificamente umani, vengono imitati e replicati con una precisione tale da rendere incerta la distinzione netta tra l'opera creativa dell'essere umano e il prodotto determinato della macchina artificiale. Turing coglie che il vero problema metodologico non sta nella macchina

⁴ A. Turing, "Computing Machinery and Intelligence," in «Mind», LIX, 236, 1950, p. 433.

in sé, ma nella nostra persistente incapacità di definire un limite chiaro e non arbitrario tra intelligenza autentica e simulazione convincente, un limite che non sia puramente pregiudiziale.

Al centro di questa discussione si colloca inevitabilmente il concetto di coscienza, un fenomeno che rimane in parte oscuro anche per la stessa indagine scientifica e che viene molte volte invocato come ultimo elemento solamente umano, elemento irriducibile che separerebbe in maniera definitiva l'essere umano dalla macchina. Uno studio recente ha rilevato l'esistenza di almeno ventinove diverse teorie scientifiche sulla coscienza umana, teorie che molte volte ricorrono persino a concetti presi in prestito dalla meccanica quantistica per tentare di spiegarne il funzionamento misterioso.⁵ Si crea così una situazione epistemologica paradossale: vogliamo spiegare un fenomeno assolutamente complicato che non comprendiamo bene usando teorie fisiche che sono esse stesse molto difficili da capire e verificare. In questa visione risulta problematico pretendere che una macchina abbia necessariamente qualcosa chiamato coscienza nel senso forte del termine, quando manca totalmente, a monte di questa richiesta, una definizione condivisa e precisa di cosa essa sia effettivamente nell'essere umano biologico. Alan Turing, suggerisce di sospendere il giudizio sulla natura della coscienza per concentrarsi invece sulla dimostrabilità del comportamento osservabile.

Un'altra obiezione su cui Turing si concentra, è quella che si potrebbe chiamare la stupidità delle macchine, riconducibile alla mancanza di competenza referenziale, cioè la capacità di associare i concetti linguistici al mondo reale. Questa critica afferma che i sistemi artificiali sono strutturalmente incapaci di attribuire un significato reale, legato concretamente al mondo, ai simboli che elaborano.⁶ Si tratterebbe, per usare una distinzione filosofica ormai classica, di una gestione puramente sintattica dei segni linguistici, senza alcun accesso genuino alla dimensione semantica del significato. Questa limitazione appare oggi evidente nel dibattito sui Large Language Models, dei sistemi capaci di creare testi coerenti e plausibili ma la cui capacità viene spesso liquidata come una mera ripetizione probabilistica di pattern statistici estratti dai dati di addestramento.⁷

Eppure, la sfida intellettuale di Turing resta necessaria e attuale: se la macchina riesce a gestire il linguaggio in modo da simulare perfettamente la comprensione del contesto, se è capace di riconoscere che la stessa parola può assumere significati opposti in base al contesto in cui si trova, la distinzione filosofica tra comprensione reale e manipolazione efficace di simboli diventa, per uno spettatore esterno che non ha accesso alla struttura interna del sistema, del tutto irrilevabile. È questa

⁵C. Koch, M. Massimini, M. Boly, G. Tononi, "Neural correlates of consciousness: progress and problems", *Nature Reviews Neuroscience*, vol. 17, 2016, pp. 307-321.

⁶J.R. Searle, "Minds, Brains, and Programs" *Behavioral and Brain Sciences*, vol. 3, n.3, 1980, pp.417-424.

⁷ L'espressione stochastic parrots è stata coniata da E. Bender et al., "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", in «*Proceedings of FAccT*», 2021, pp. 610-623.

indistinguibilità osservabile che conta, secondo Turing, non qualche ipotetica essenza metafisica nascosta.⁸

Ci troviamo quindi davanti a due oscurità che si specchiano: quella dei processi cognitivi umani, ancora non totalmente compresi dalla scienza, e quella dei sistemi algoritmici, le cui cosiddette scatole nere rendono molto difficile tracciare con precisione il percorso interno che porta ad un determinato output. Se accettiamo questo presupposto epistemologico, dobbiamo anche accettare un effetto filosoficamente rilevante: Il successo di sistemi artificiali nel riprodurre comportamenti linguistici complessi ci pone davanti ad un bivio interpretativo. O abbiamo finalmente compreso le leggi fondamentali che regolano il pensiero e siamo riusciti a codificarle in modo computazionale, oppure la riproduzione tecnica ha raggiunto un tale livello di sofisticazione da rendere la realtà propriamente umana una nozione quasi antico, in quanto superato dalla sua stessa immagine riflessa nella macchina.

1.2 – ‘‘Macchine calcolatrici e intelligenza’’ di Alan Turing.

Nel saggio dal 1950. ‘‘Computing Machinery and Intelligence’’ Alan Turing delinea quella che può essere considerata l’architettura filosofica fondamentale dell’intelligenza artificiale come disciplina scientifica. Il testo apre una discussione filosofica sul quesito se le macchine possano pensare, muovendo dalla consapevolezza metodologica che tale domanda richiederebbe, per essere affrontata in modo rigoroso, una definizione preliminare condivisa di concetti estremamente sfuggenti come quelli di pensiero e di macchina.⁹ La strategia argomentativa di Alan Turing, estremamente contemporanea nella sua concretezza, consiste nello spostare la questione metafisica tradizionale con un test operativo e comportamentale: il cosiddetto Gioco dell’Imitazione, che diventerà poi famoso come Test di Turing. Tale dispositivo concettuale non è un semplice esperimento mentale tra i tanti, ma rappresenta un vero e proprio ribaltamento di visione filosofica sulla natura dell’intelligenza.

Nel gioco originale così come Turing lo descrive, i partecipanti sono tre; un uomo che chiamiamo B, una donna che chiameremo A, e un interrogante che chiameremo C. Quest’ultimo è separato fisicamente dagli altri due e deve identificare il genere basandosi solamente sulle risposte scritte che riceve alle sue domande. Mentre l’uomo B ha il compito di aiutare l’interrogante nella sua identificazione, la donna A deve invece cercare di ingannarlo facendosi passare per l’uomo, creando così un interessante paradosso epistemologico in cui la verità deve essere separata da una finzione

⁸ A. Turing, ‘‘Computing Machinery and Intelligence’’, cit., p. 446.

⁹ Ivi, p. 433.

ben costruita e plausibile.¹⁰ La vera svolta epistemologica di Turing avviene però nel momento in cui egli introduce la variante tecnologica cruciale: cosa accadrebbe se una macchina prendesse il posto di A? se l'interrogante desse la risposta errata con la stessa frequenza statistica con cui la darebbe in un gioco tra soli uomini, allora la questione originale, se le macchine possano pensare, verrebbe di fatto risolta sul piano pratico e operativo.

In questo preciso passaggio concettuale Turing compie un'operazione filosofica di smontaggio della soggettività: riorienta il terreno di indagine dall'essenza metafisica al comportamento osservabile; quindi, dal che cosa è al che cosa fa. È interessante notare come Turing sottolinei esplicitamente che il gioco traccia una linea di separazione abbastanza netta tra le capacità fisiche e quelle intellettuali dell'uomo.¹¹ In questo modo la tecnica viene deliberatamente isolata nella sua dimensione puramente logica e funzionale, svincolata da ogni necessità di realizzazione corporea o biologica scientifica. Per Alan Turing la capacità degli scienziati di riprodurre in modo artificiale l'aspetto fisico dell'essere umano è totalmente irrilevante filosoficamente. Ciò che per lui conta davvero è la performance intellettuale osservabile attraverso il linguaggio.

Il metodo domanda-risposta si dimostra infatti uno strumento estremamente forte, in grado di esplorare, come Turing stesso afferma, quasi ogni campo della conoscenza umana: dalla sensibilità artistica necessaria per comporre un sonetto alla pura razionalità logica richiesta per giocare a scacchi.¹² Il gioco non vuole penalizzare la macchina per i suoi limiti fisici evidenti, né tantomeno l'uomo per la sua relativa lentezza nel calcolo numerico, ma istituisce invece un terreno di confronto simbolico e linguistico che Turing identifica come il dominio reale e proprio del pensiero in quanto tale. Egli arriva a sostenere una tesi molto potente: se una macchina è costruita in modo da giocare il gioco dell'imitazione in modo soddisfacente, superando il test con una certa regolarità, non abbiamo più nessun tipo di bisogno epistemologico di farci domande sulla natura del suo processo interno o sulla sua eventuale coscienza, quel che non è discernibile nel comportamento osservabile non possiede, di fatto e di diritto, nessuna rilevanza operativa.

Avendo fissato questo criterio comportamentale, Turing dedica la parte centrale e più corposa del suo saggio ad una confutazione sistematica delle obiezioni più comuni e resistenti, che ancora oggi dominano il dibattito filosofico sull'IA. La prima che esamina è l'obiezione teologica, secondo cui Dio avrebbe dotato di anima solo gli esseri umani e non le macchine, rendendo quindi impossibile per principio che una macchina possa veramente pensare. Turing smonta questa obiezione, affermando che negare a Dio la possibilità di conferire un'anima ad una macchina, significherebbe

¹⁰ Ivi, pp. 433-434.

¹¹ Ivi, p. 434.

¹² Ivi, p. 435.

paradossalmente limitarne l'onnipotenza divina, che è invece per definizione illimitata.¹³ In questo modo l'argomentazione teologico si mostra contraddittorio dalle sue stesse premesse.

Allo stesso modo è interessante la critica che Turing rivolge all'obiezione della testa nella sabbia, in cui egli vede non una posizione filosofica argomentata razionalmente, ma un sentimento di paura culturale dell'uomo davanti alla possibilità reale di perdere il proprio primato cognitivo nel creato. È la resistenza emotiva tipica di chi non vuole accettare che qualcosa di fondamentale nella nostra auto-comprensione come specie possa essere messo in discussione dal progresso tecnologico¹⁴. Ma la questione veramente importante della disputa filosofica è l'argomento dell'autocoscienza, portato avanti dal neurochirurgo Geoffrey Jefferson in una conferenza del 1949, secondo cui una macchina non potrebbe mai essere considerata puramente intelligente fino a quando non senta veramente le proprie creazioni, fino a quando non sia conscia in prima persona di quello che lei stessa crea.¹⁵

La risposta di Alan Turing a questa obiezione è molto sottile. Egli fa notare che anche la nostra attribuzione della coscienza agli altri esseri umani è necessariamente basata su un'inferenza basata sul comportamento esterno osservabile, non su un accesso diretto e privilegiato alla loro interiorità soggettiva. Non abbiamo mai accesso diretto alla coscienza degli altri, cerchiamo di coglierla dai comportamenti, dal linguaggio, dalle azioni. Se quindi non pretendiamo una prova introspettiva diretta di coscienza dai nostri simili esseri umani per poterli considerare pensanti e coscienti, allora pretenderla dalla macchina significa adottare, forse senza rendersene conto, una posizione solipsistica insostenibile secondo cui solo io so con certezza di essere cosciente¹⁶. L'asimmetria di criteri applicati sarebbe ingiustificata e metodologicamente arbitraria.

Infine, c'è l'obiezione di Lady Lovelace, matematica britannica dell'Ottocento, affermava che una macchina può fare solo ciò che le ordiniamo di fare e quindi non può mai sorprenderci o mostrarsi creativa in sé stessa. Turing controbatte questa obiezione in due modi complementari. Prima di tutto ridefinisce il concetto stesso di creatività umana, osservando con precisione che anche il pensiero umano è molte volte l'effetto indiretto di un seme gettato dall'istruzione ricevuta, dall'educazione, e quindi non nasce mai totalmente dal nulla in modo spontaneo. In secondo luogo, Turing osserva che le macchine complesse possono effettivamente prendere alla sprovvista il loro stesso programmatore, proprio a causa dell'impossibilità pratica per l'essere umano di prevedere ogni singola conseguenza logica di un sistema computazionale che possiede miliardi di stati possibili e interazioni complesse tra le sue parti. Con questa analisi Alan Turing, non si limita solamente a prendere le difese di

¹³ Ivi, p. 434.

¹⁴ Ivi, p.435.

¹⁵ Ivi, p. 444.

¹⁶ A. Turing, "Computing Machinery and Intelligence", cit., p. 446.

un'invenzione tecnica particolare, ma definisce e fonda una vera e propria filosofia della tecnica in cui l'intelligenza è concepita come un fenomeno concreto, osservabile, legato essenzialmente alla capacità di agire efficacemente nel mondo del linguaggio e del simbolo.¹⁷

1.3 – Turing e l'eresia del controllo delegato

Successivamente al saggio del 1950 sul Gioco dell'Imitazione, il pensiero di Turing si evolve ancora di più. Tale evoluzione è documentata in un intervento del 1951, denominato "Intelligent Machinery, A Heretical Theory", un saggio nato inizialmente per un programma radiofonico della BBC e poi rielaborato in forma scritta più ampia.¹⁸ La scelta dell'aggettivo eretica nel titolo non è casuale o fine a sé stessa. Turing è completamente consapevole di voler sfidare uno dei dogmi più radicati e resistenti della cultura occidentale moderna, ovvero l'idea che l'essere umano debba sempre e necessariamente mantenere un controllo assoluto su ogni sua creazione tecnica. In questa sede il suo obiettivo non è più solamente rispondere alla questione astratta sulla possibilità del pensiero artificiale, ma scardinare esplicitamente il pregiudizio secondo il quale non si può e non si deve fare in modo che una macchina pensi per noi, al posto nostro.

Contro questa proibizione culturale, che Turing considera un limite ideologico alla libera ricerca scientifica più che una posizione filosofica argomentata razionalmente, viene articolato un pensiero caratterizzato da una fiducia quasi illimitata nel metodo ingegneristico e nelle sue potenzialità. La sua risposta si articola lungo due strade molto interessanti e dense. Prima di tutto egli approfondisce e sviluppa la nozione di macchina che apprende, immaginando un vero e proprio processo educativo simile a quello umano. Alan Turing non pensa più semplicemente a una programmazione rigida e predeterminata in ogni dettaglio, ma introduce il concetto rivoluzionario di "Unorganized Machine", un sistema elementare e inizialmente poco strutturato che attraverso l'esperienza pragmatica e un meccanismo di risposte continuo basato su punizioni e ricompense, quello che lui chiama il "Pleasure-pain system", possa progressivamente auto-organizzarsi fino a gestire situazioni impreviste e non anticipate dal programmatore.

È interessante notare come Alan Turing faccia riferimento in modo esplicito all'analogia pedagogica umana: immagina per la macchina uno sviluppo cognitivo del tutto simile a quello di un bambino, che necessita di un educatore attento per selezionare e ordinare gli stimoli del mondo esterno, che

¹⁷ A. Turing, "Computing Machinery and Intelligence", cit., p. 450.

¹⁸ A.M Turing, "Intelligent Machinery, A Heretical Theory" 1951, in B.J Copeland (a cura di), The Essential Turing, Oxford University Press, Oxford 2004, p.465

altrimenti sarebbero caotici e incomprensibili. In questa prospettiva innovativa il comportamento non previsto anticipatamente da un computer non viene più colto come un errore di sistema, ma come il segno positivo di un apprendimento autentico, un risultato che, pur non essendo stato anticipato nei minimi dettagli dal creatore umano, rientra comunque in un ambito di progresso tecnico desiderato.²⁰ La macchina diviene qualcosa di simile ad un allievo che sorprende il maestro con una soluzione originale a un problema.

All'interno del saggio, Turing arriva perfino ad ipotizzare, che una volta superato un determinato livello critico di intelligenza e di capacità autonoma di interazione simbolica, le macchine potrebbero effettivamente e concretamente prendere il controllo di molti processi decisionali umani.²¹ Egli ci invita a considerare questa possibilità inquietante con lo stesso distacco emotivo ma temporalmente molto lontano, come il raffreddamento progressivo del sole tra miliardi di anni. È qui che emerge con chiarezza quella che potremmo definire una vera e propria filosofia del controllo delegato. Per Alan Turing il successo ultimo e più completo della tecnica non sta nel comando continuo dell'uomo sulla macchina, ma nella capacità di progettare entità così affidabili da rendere la delega completa del controllo non solo possibile in chiave tecnico, ma anche desiderabile.

L'incontrollabilità del fine, in questa visione ottimista, smette di essere percepita come un rischio da evitare e diventa il simbolo e la prova del trionfo definitivo dell'ingegno umano. Abbiamo creato qualcosa di così complesso e potente da non necessitare più di una sorveglianza continua e attenta da parte nostra.

Turing traccia così un orizzonte concettuale in cui il progresso tecnico e scientifico coincide paradossalmente, con la liberazione progressiva dell'uomo attraverso la sua stessa, volontaria e consapevole, sostituzione negli ambiti del controllo decisionale e dell'organizzazione razionale dei processi complessi.

1.4 – La dissoluzione del mistero: il funzionalismo e la riduzione logica dell'essere umano.

Indagando ulteriormente il progetto di Turing, ci si accorge che la sua proposta non si limita ad un'innovazione nel campo della tecnica, ma configura una vera e reale operazione di smontaggio dell'essere umano. Attraverso la sostituzione dei termini teorici con definizioni puramente operative, si assiste alla dissoluzione di quel mistero della coscienza che per anni ha rappresentato l'uomo l'ultima barriera dell'eccezionalismo umano. È proprio qui si intravede il centro di quello che la filosofia successiva denomina funzionalismo; l'idea, cioè, che gli stati mentali non debbano essere

²⁰ Ivi, p. 471.

²¹ Ivi, p. 475.

definiti dalla loro intrinsecità, dalla specifica sostanzialità, ma solamente dalla funzione che svolgono “all’interno di un sistema logico e casuale.”²²

Questa visione porta ad una demitizzazione dell’interiorità che potremmo definire quasi radicale. Se accettiamo il criterio del Gioco dell’imitazione, allora dobbiamo anche accettare che l’introspezione e tutto ciò che è soggettivo, perda la sua posizione privilegiata. Alan Turing suggerisce, in maniera implicita, che non vi sia alcuna differenza tra un sistema che comprende veramente e un sistema che agisce “come se” comprendesse, a patto che i risultati non siano distinguibili²³. È un approccio che punta a ricondurre l’essere umano entro i limiti del calcolabile, eliminando ogni residuo di trascendenza. La concezione di uomo cambia radicalmente, viene riletto come una sorta di macchina a strati discreti. Una macchina sicuramente complessa, ma pur sempre finita e comprensibile nelle sue dinamiche di input e output.

Interessante è come questa visione trasformi la mente in un oggetto d’indagine scientifica uguale di qualsiasi altro apparato meccanico. Se da una parte tale dissoluzione del mistero alimenta un ottimismo senza precedenti cioè l’idea che ogni capacità dell’uomo sia replicabile, dall’altro prepara il terreno per quella crisi del soggetto che emergerà nel pensiero di Anders. Riducendo l’essere umano ad un insieme di funzioni, ad un qualcosa di totalmente analizzabile e misurabile²⁴.

In questo modo anche la competenza referenziale in quest’ottica viene ridimensionata; se il significato può essere ridotto ad una complessa rete di relazioni e connessioni logiche, allora anche il linguaggio cessa di essere un veicolo di verità profonde per diventare un sofisticato sistema di manipolazione simbolica. Il funzionalismo di Alan Turing, in conclusione, decreta la fine dell’antropocentrismo classico. Al suo posto resta un modello in cui l’essere umano e la macchina sono colti come istanze diverse della sua stessa capacità logica universale, pareggiando ogni differenza che non sia meramente computazionale.

²² Cfr. N. Block, “What is Functionalism?” In “Readings in Philosophy of Psychology», Harvard University Press, Cambridge 1980.”

²³ A. Turing, “Macchine calcolatrici e intelligenza”, cit., p. 132.

²⁴ M. Minsky, “Step Toward Artificial Intelligence”, Proceedings of the IRE, vol. 49, n.1, 1961, pp. 8-30.

1.5 – La smaterializzazione del soggetto: verso un'intelligenza senza corpo.

Turing definisce e coglie l'intelligenza come una proprietà universale che non ha alcun debito verso la "macchina proteica" del cervello umano²⁵. In questo modo, la mente viene isolata dal corpo, dai sensi e dalla dimensione organica, per essere tradotta in un sistema di simboli e stati completamente manipolabili.

Ciò che fuoriesce è la figura di un soggetto disincarnato. Se il "Gioco dell'imitazione" richiede in modo esplicito di nascondere il corpo dietro ad uno schermo, è perché Turing crede che il corpo sia poco più di un rumore di fondo, un elemento accidentale che non aggiunge valore alla veridicità del pensiero. Quella linea abbastanza netta tra ciò che fisico e ciò che è intellettuale nel saggio del 1950, diventa in questo modo il manifesto di una nuova epoca²⁶. L'identità non è più ancorata alla presenza biologica, ma alla possibilità di creare flussi di informazioni coerenti. L'intelligenza viene tolta al tempo del decadimento, per essere trasferita nel dominio dell'astrazione logica, dove tutto può essere eterno e replicabile.

Ciò nonostante, proprio in questa smaterializzazione si coglie la massima frizione con la nostra esperienza vissuta. Togliendo il corpo, Alan Turing, elimina anche la vulnerabilità e la finitudine. L'ottimismo presente in tutto il suo progetto sta proprio nella convinzione che tale "liberazione" dalla materia sia un progresso auspicabile. Una macchina che riflette senza il peso del corpo è un'entità che può operare con una velocità e una precisione che noi non potremo mai raggiungere. È un'intelligenza pura, che non conosce quasi nessuna fragilità e che, proprio per questo, può essere potenziata all'infinito.

È chiaro quindi, come Turing abbia ridefinito la razionalità universale svincolandola dai limiti della nostra specie. Ma se la l'intelligenza può essere smaterializzata e resa perfetta in un apparato tecnico, cosa ne è dell'uomo che rimane, invece, prigioniero della sua corporeità antiquata e difettosa?²⁷

La riflessione di Alan Turing scava in qualche modo, un solco incolmabile tra la perfezione del prodotto la debolezza del produttore. Ed è qui che l'orgoglio del progetto si trasforma in quella vergogna di fronte all'oggetto che Anders analizzerà.

²⁵ Cfr. B. J. Copeland, *The Essential Turing*, Oxford University Press, Oxford 2004, p. 441.

²⁶ A. Turing, "Macchine calcolatrici e intelligenza", cit., p. 128.

²⁷ G. Anders, "L'uomo è antiquato". Vol. I: *Sull'anima nell'epoca della seconda rivoluzione industriale*, Bollati Boringhieri, Torino 2002, pp. 23-25.

Capitolo 2- Gunther Anders, la tecnica come minaccia esistenziale.

2.1 – La condizione umana nell'epoca tecnica: genesi e fenomenologia della “Vergogna Prometeica”

Il passaggio dal pensiero di Turing alla diagnosi impetuosa di Gunther Anders segna quello che può essere definito il trauma della materialità nell'epoca della tecnica moderna. Con Anders ci troviamo di fronte ad una riflessione in qualche modo speculare: la riscoperta dolorosa della corporeità non più come luogo della soggettività autentica, ma come peso esistenziale che ci rende strutturalmente inferiori rispetto ai nostri stessi prodotti tecnici. La nozione centrale che inaugura questa fase di analisi è la “Prometheische Scham”, la vergogna prometeica, una categoria che Anders sviluppa nella sua opera principale “L'uomo è antiquato” pubblicata in due volumi nel 1956 e 1980.²⁸

La vergogna prometeica non è semplicemente un sentimento psicologico individuale o una forma di complesso di inferiorità patologico, ma rappresenta invece una condizione antropologica strutturale nell'epoca contemporanea, una trasformazione radicale nel modo in cui l'uomo percepisce sé stesso e il proprio rapporto con il mondo tecnico che lo circonda. Anders descrive questo fenomeno con una formulazione provocatoria: l'uomo si vergogna di essere nato invece di essere stato fatto. Si vergogna cioè della propria origine biologica, imperfetta, non progettata, davanti alla perfezione calcolata, all'affidabilità dei prodotti tecnici²⁹. Mentre le macchine sono create con attenzione, ottimizzate per lo scopo, create secondo criteri di efficienza massima. L'essere umano biologico è il risultato di un processo evolutivo cieco, pieno di ridondanze, imperfezioni e fragilità strutturali.

È fondamentale evidenziare che questa vergogna non nasce da una reale inferiorità oggettiva dell'uomo rispetto ad una macchina in senso assoluto. Ma da una trasformazione nel criterio stesso con cui valutiamo l'eccellenza e il valore. Nell'epoca della tecnica matura, il criterio dominante diventa l'efficienza, la precisione, la velocità. E rispetto a tali criteri, che sono ovviamente criteri tecnici e non antropologici, l'uomo risulta effettivamente inadeguato. Un computer non si stanca, non ha bisogno di riposare. La macchina rappresenta quindi una sorta di ideale irraggiungibile di perfezione funzionale, e questa irraggiungibilità crea nell'uomo un sentimento di inadeguatezza esistenziale profonda.

²⁸ G. Anders, “L'uomo è antiquato, vol. I: Sull'anima nell'epoca della seconda rivoluzione industriale”, tr. it. L. Dallapiccola, Bollati Boringhieri, Torino 2002 (ed. or. Die Antiquiertheit des Menschen, C.H. Beck, München 1956).

²⁹ Ivi, p. 28

Anders identifica questa condizione come il nucleo del dislivello prometeico, espressione che indica lo scarto crescente e incolmabile tra ciò che l'essere umano è capace di creare tecnicamente e ciò che è in grado di immaginare moralmente, di comprendere nelle sue conseguenze, di controllare responsabilmente. L'essere umano prometeico dell'epoca contemporanea ha rubato il fuoco agli dei, ha acquisito poteri tecnici enormi, ma è rimasto per altri versi molto limitato: la sua immaginazione morale, la sua capacità di sentire emotivamente la portata delle proprie azioni, la sua struttura psicologica sono rimaste quelle di un primate evoluto in piccoli gruppi di cacciatori-raccoglitori.³⁰ Possiamo costruire armi in grado di annientare l'intera civiltà, ma non riusciamo veramente a sentire, a percepire emotivamente cosa vuol dire questa possibilità.

L'effetto più drammatico di questa situazione è quella che Anders chiama l'obsolescenza dell'essere umano. Non si tratta di una metafora retorica ma di una diagnosi filosofica precisa: l'essere umano sta diventando oggettivamente antiquato, superato, reso obsoleto dalle sue stesse creazioni. Non solo nel senso banale che le macchine sono più veloci o precise in certi compiti specifici, ma nel senso più profondo che il mondo tecnico che abbiamo costruito non è più fatto a misura umana, non rispecchia più la nostra scala percettiva, i nostri tempi biologici, i nostri limiti cognitivi. Viviamo in un ambiente tecnico che ci supera strutturalmente, e questa sproporzione genera un'alienazione ontologica molto più profonda; l'alienazione dell'essere umano rispetto alla propria stessa epoca storica.

2.2 – Il dislivello prometeico: produrre senza comprendere.

La nozione di dislivello prometeico, che abbiamo già introdotto, merita un'analisi molto più approfondita poiché rappresenta il centro teorico della filosofia di Anders. Questo dislivello indica letteralmente un dislivello, un gradiente, uno scarto verticale tra due livelli. Ma quali sono esattamente questi due livelli? Da una parte c'è la capacità produttiva dell'essere umano, ciò che siamo tecnicamente in grado di fare, i sistemi che siamo capaci di costruire, le forze che siamo in grado di scatenare. Dall'altra parte c'è la capacità immaginativa, la facoltà di rappresentarci mentalmente gli effetti di ciò che produciamo, di sentire emotivamente, di assumerci la responsabilità morale per esse.³¹

Questo dislivello non è una condizione transitoria destinata ad essere superata, ma anzi strutturalmente tende ad aumentare progressivamente con l'accelerazione dello sviluppo tecnico. Ogni nuova generazione di tecnologie amplia ulteriormente lo scarto tra potere e comprensione.

³⁰ Ivi, pp. 23-30

³¹ Ivi, pp. 31-40

Possiamo creare effetti su scale temporali e spaziali che superano completamente la nostra capacità biologica di percezione diretta. Possiamo influenzare sistemi così complessi che nessuna mente umana individuale può comprenderli nella loro totalità. Possiamo scatenare processi irreversibili i cui effetti si dispiegheranno su generazioni future che non possiamo neanche immaginare concretamente. Anders mostra questo concetto con molti esempi tratti dalla realtà del suo tempo, ma quello più potente e ricorrente riguarda la bomba atomica, che per lui rappresenta il caso concreto del dislivello prometeico portato alle sue conseguenze estremo. Gli scienziati del Progetto Manhattan erano perfettamente capaci, sul piano tecnico e scientifico, di calcolare con precisione la potenza dell'esplosione nucleare, la quantità di energia liberata. Ma non erano in grado, sul piano immaginativo ed emotivo, di sentire davvero cosa significasse questa potenza. Non potevano cogliere concretamente, con partecipazione emotiva autentica, cosa volesse dire eliminare istantaneamente decine di migliaia di vite umane, condannare altre decine di migliaia a una morte lenta per radiazioni.³²

Questa incapacità non è una mancanza morale individuale degli scienziati coinvolti, molti dei quali erano persone moralmente sensibili che in seguito mostrarono profondi dubbi e rimorsi. È una limitazione strutturale della condizione umana davanti a fenomeni che superano la scala dell'esperienza biologica possibile. Possiamo immaginare la sofferenza di una persona, forse di dieci persone, forse anche di cento. Ma oltre ad una certa soglia i numeri diventano astrazioni statistiche che non toccano più il sentimento morale. Questo fenomeno Anders lo denomina con un termine molto efficace: "Gefühlslosigkeit", letteralmente assenza di sentimento, ma nel senso specifico di incapacità di sentire emotivamente la portata di ciò che si produce tecnicamente. Non è indifferenza morale consapevole, ma impossibilità strutturale di tradurre la conoscenza astratta in percezione emotiva pragmatica. Sappiamo intellettualmente che una bomba nucleare uccide centomila persone, ma non riusciamo a sentire questo dato ugualmente al modo in cui sentiremmo la morte di una persona cara. E questa sproporzione tra sapere e sentire è il cuore del dislivello prometeico.³³

Ma il dislivello non riguarda solo la scala quantitativa degli effetti. Fa riferimento anche alla complessità qualitativa dei sistemi tecnici oggi. Un sistema tecnico sufficientemente complesso, come una centrale nucleare, una rete elettrica nazionale, un moderno sistema di IA, non è più comprensibile nella sua totalità da nessun singolo individuo. Ogni esperto comprende la propria parte, il proprio dominio di competenza, ma nessuno ha una visione d'insieme completa. Il sistema nel suo complesso è diventato più intelligente, o almeno più complesso, di qualsiasi intelligenza umana individuale che contribuisca a mantenerlo.

³² G. Anders, "L'uomo è antiquato" Vol. II: Sulla distruzione della vita nell'epoca della terza rivoluzione industriale, Bollati Boringhieri, Torino 2003, pp. 233-270.

³³ Ivi, pp. 261-275.

2.3 – Il mondo come fantasma: la mediazione tecnica dell'esperienza.

Un altro aspetto centrale della filosofia di Anders, strettamente collegata al dislivello prometeico ma dotata di una sua specificità, è la distinzione tra il mondano e il fantasma. Questa coppia concettuale, che Anders sviluppa analizzando fenomeni come la televisione e la fotografia ma si mostra straordinariamente attuale per comprendere l'esperienza mediata digitalmente, descrive la trasformazione radicale nel modo in cui facciamo esperienza della realtà nell'epoca della riproduzione tecnica.³⁴

Il mondano, nelle parole di Anders, sottolinea la realtà così come la incontriamo attraverso l'esperienza corporea diretta, non mediata tecnicamente. È il mondo nella sua resistenza fisica, nella sua materialità concreta, nella sua presenza immediata ai nostri sensi biologici. Quando tocchiamo un oggetto e sentiamo la sua superficie ruvida o liscia, quando camminiamo e sentiamo la fatica nei muscoli, quando abbiamo fame o freddo, quando percepiamo il dolore: tutto questo appartiene al registro del mondano. È un'esperienza che non può essere perfettamente tradotta in segni, in rappresentazioni, in dati. Ha una qualità fenomenologica irriducibile che si dà solo nell'essere-lì, nella presenza corporea concreta.

Il fantasma è invece l'immagine tecnica, la rappresentazione mediata, il dato digitale che sostituisce progressivamente l'esperienza diretta. La fotografia, il film, la trasmissione televisiva, e oggi lo schermo digitale, sono tutti fantasmi in questo senso tecnico. Non sono illusioni nel senso di falsità, possono anzi essere molto accurati e informativi. Ma sono rappresentazioni che ci raggiungono attraverso dispositivi tecnici, già processate, già selezionate, già filtrate secondo criteri che non controlliamo direttamente. E questa mediazione non è neutrale, trasforma in modo qualitativo la natura dell'esperienza stessa.³⁵

Anders coglie un fenomeno che oggi è divenuto ancora più pervasivo; viviamo sempre più nel mondo dei fantasmi e sempre meno in quello mondano. Gran parte della nostra conoscenza del mondo non deriva da esperienza diretta ma da immagini tecnicamente mediate. Sappiamo com'è fatto un elefante non perché ne abbiamo mai visto uno dal vivo, ma perché l'abbiamo visto su fotografie o video. Sappiamo cosa sta accadendo dall'altra parte del mondo perché lo vediamo dallo schermo. Conosciamo luoghi, persone, eventi attraverso la loro rappresentazione fantasmatica, non attraverso l'incontro mondano.

Ma l'effetto più problematico di questa trasformazione non è semplicemente che riceviamo informazioni mediate, il che di per sé potrebbe essere colto come un arricchimento delle nostre

³⁴ G. Anders, "L'uomo è antiquato", vol. I, pp. 115-140 – "Il mondo come fantasma e come matrice analizza gli effetti della televisione."

³⁵ Ivi, pp. 120-125.

possibilità conoscitive. Il problema è che progressivamente perdiamo la capacità di distinguere qualitativamente tra mondano e fantasma, tra esperienza diretta e rappresentazione mediata. E ancora più grave: cominciamo a preferire il fantasma al mondano.

La realtà mondana è irripetibile, irreversibile, non negoziabile.

Questa preferenza per il fantasma crea quello che Anders chiama un processo di derealizzazione del mondo. La realtà concreta, corporea, comincia a sembrare meno importante, meno degna di attenzione rispetto alla sua rappresentazione tecnica. E questo ha effetti morali devastanti. Se un evento accade ma non viene filmato, se non diventa immagine, sembra non essere davvero accaduto. Se una sofferenza non viene mediatizzata, se non diventa contenuto, perde rilevanza. Il mondo mondano viene a mano a mano svuotato di significato e valore, che migrano totalmente nel regno dei fantasmi.

2.4– La delega della responsabilità in ‘Noi figli di Eichmann’.

‘Noi figli di Eichmann’ (1964), è un testo nato da uno scambio epistolare con Claude Eatherly (il pilota che diede il via libera al bombardamento di Hiroshima). Anders analizza la figura di Adolf Eichmann (burocrate che organizzò i trasporti per la soluzione finale), non come un’eccezione mostruosa, ma come il prototipo dell’uomo nel tempo della tecnica.³⁶

Il concetto centra nel saggio è la ‘Banalità del male Tecnologico’, che Anders interpreta attraverso le lente dell’apparato. Eichmann, al processo si difese affermando di essere una sorta di ingranaggio di un meccanismo più grande, un esecutore di ordini che non aveva nulla contro le vittime, ma che semplicemente faceva il suo lavoro. Anders coglie che questa non è solo una scusa processuale, ma soprattutto, la descrizione fedele di una nuova condizione umana; nell’epoca della tecnica, l’azione è stata sostituita dal lavorare all’interno dell’apparato.³⁷

Anders, sostiene che l’inserimento dell’uomo in un apparato tecnico (che sia una burocrazia statale o un sistema di algoritmi di IA), produce una scissione tra l’azione e l’immaginazione degli effetti. Se l’azione è parcellizzata (se un uomo si limita a programmare un modulo, un altro a raccogliere dati, e un terzo a lanciare l’algoritmo), nessuno si sentirebbe responsabile del risultato finale, anche se

³⁶ G. Anders, ‘Noi figli di Eichmann’, Giuntina, Firenze 1995 (ed. or. 1964).

³⁷ Ivi, pp. 34-40. Anders parla della trasformazione dell’azione in “prestazione” all’interno di un apparato che rende invisibile il fine ultimo.

questo comporta discriminazione di massa o catastrofi economiche. È quella che in letteratura accademica viene definita “Irresponsabilità per eccesso di mediazione”.³⁸

Nel mondo tecnico, si nota come l'apparato agisca da vero e proprio schermo morale. La delega alla macchina o alla procedura arresta la coscienza perché l'uomo non coglie più il nesso causale tra il suo gesto (premere un tasto, scrivere un codice) e la sofferenza dell'altro. Siamo diventati “analfabeti emotivi” proprio perché la tecnica è troppo veloce e vasta per la limitata capacità di sentire dell'essere umano.³⁹ Anders definisce questa condizione come “nichilismo della tecnica”: non è l'uomo voglia il male, è che il male si è trasformato in un prodotto collaterale di un funzionamento efficiente a cui l'uomo non sa più opporsi.

Questa riflessione ci porta al nucleo del dislivello prometeico applicato all'etica. L'umanità edifica macchine che prendono decisioni morali, ma è più capace di assumersi la responsabilità di quelle decisioni. La frase “è stato l'algoritmo” è la versione contemporanea del “seguivo gli ordini” di Eichmann. In tutti e due i casi, l'uomo si spoglia della sua umanità per farsi cosa tra le cose, rinunciando al giudizio in cambio della sicurezza del protocollo. La delega di Alan Turing, dal punto di vista di Anders, si mostra così essere la forma più originale di suicidio morale; l'uomo cessa di essere antiquato solo nel momento in cui smette di essere uomo e diventa una funzione perfetta dell'apparato.⁴⁰

2.5 – La smaterializzazione dell'esperienza: il mondo come fantasma e l'uomo come pezzo di ricambio.

Il dislivello prometeico non si limita a generare un sentimento di inferiorità, ma spinge l'uomo verso una vera e propria strategia di sopravvivenza ontologica: la trasformazione di sé in “pezzo di ricambio”⁴¹.

La riflessione di Anders è di una sottigliezza estrema. Egli osserva che, in un mondo dominato da apparati tecnici perfetti, seriali e immortali, l'uomo soffre la propria unicità. Se nella tradizione filosofica e umanistica l'essere unici era un pregio, nel sistema tecnico l'unicità è un difetto, poiché ciò che è unico non è sostituibile e, dunque, inefficiente. La vergogna prometeica cambia qui in un

³⁸ Cfr. Z. Bauman, “Modernità e Olocausto”, Il Mulino, Bologna 1992, dove l'autore riprende esplicitamente le tesi di Anders sulla distanza morale creata dalla burocrazia tecnica.

³⁹ C. Dries, “Günther Anders's Philosophical Anthropology: Existential Threat and the Question of Praxis”, Thesis Eleven, vol. 120, n.1, 2014, pp. 27-47. Analizza il concetto andersiano di azione “fantasmatica” e deresponsabilizzazione.

⁴⁰ G. Anders, “Noi figli di Eichmann”, cit., p. 52.

⁴¹ G. Anders, “L'uomo è antiquato”. Vol. I: Sull'anima nell'epoca della seconda rivoluzione industriale, Bollati Boringhieri, Torino 2002, p. 85.

desiderio di standardizzazione. L'essere umano inizia a modellare la propria vita, i propri ritmi e persino i propri sogni sull'immagine della macchina, cercando di togliere tutto ciò che in lui è resistente, lento o imprevedibile, e quindi tipicamente umano.⁴²

Andando a fondo questa tesi, emerge chiaramente come l'uomo cerchi di farsi cosa per sfuggire al dolore della propria finitudine. È quella che potremmo definire auto-reificazione. Se Alan Turing coglieva nella macchina la possibilità di espandere le capacità umane, Anders inverte la visione; è 'essere umano che si contrae per adattarsi alla forma della macchina stessa. Non siamo più noi ad utilizzare l'apparato, ma è l'apparato che ci usa come suoi componenti organici. L'uomo, quindi, diventa un "pezzo di ricambio", non solo nel processo produttivo, ma nella struttura stessa della realtà sociale e conoscitiva.⁴³

Questa rinuncia alla soggettività è il prezzo da pagare che l'uomo paga per essere ammesso nel mondo tecnico. L'obsolescenza dell'uomo si realizza dunque nel momento in cui egli smette di essere un soggetto unico per diventare una funzione intercambiabile.

In questo cambiamento, il cerchio della tecnica si chiude: l'uomo ha creato della macchina a sua immagine e somiglianza per poi scoprire, con vergogna, che la nostra somiglianza era difettosa, e ha iniziato a ricostruire noi stessi a immagine e somiglianza delle macchine.⁴⁴

2.6 – L'inversione della libertà: lo scarto tra rappresentazione e produzione.

Un altro aspetto fondamentale del pensiero di Anders riguarda quello che potremmo chiamare lo "scarto temporale" tra il momento della produzione tecnica e quello della comprensione dei suoi effetti. Anders coglie che mentre la tecnica avanza con una velocità esponenziale, la capacità dell'essere umano, di immaginare e rappresentarsi le conseguenze di tale avanzamento rimane ancorata ai ritmi della sua natura biologica. Questo sfasamento temporale crea una condizione paradossale: l'uomo è sempre in ritardo rispetto a sé stesso.

Quando finalmente comprende le implicazioni di una tecnologia, quest'ultima è già stata sostituita da una successiva ancora più forte e ancora più incomprensibile. È come se corresse dietro alla sua ombra senza mai raggiungerla. La libertà, in questo contesto, subisce un'inversione profonda: l'essere

⁴² Ivi, pp. 90-95.

⁴³ P. Portinaro, "Günther Anders: un pensiero senza dimora", prefazione a G. Anders, *Essere o non essere. Diario di Hiroshima e Nagasaki*, Einaudi, Torino 1961, pp. V-XXX. Per il tema specifico del dislivello prometeico si rinvia anche a Id, "Il principio di disperazione. Tre studi su Anders". Bollati Boringhieri, Torino, 2003.

⁴⁴ G. Anders, "L'uomo è antiquato. Vol. I", cit. p. 102.

umano non è più libero di utilizzare la tecnica per i suoi stessi fini, ma è costretto a rincorrere gli effetti imprevisi che essa crea tentando disperatamente di porvi rimedio con altra tecnica.

Anders chiama questo meccanismo “la tirannia del già fatto”⁴⁶. Una volta che una tecnologia esiste, una bomba nucleare, un algoritmo di riconoscimento facciale, un sistema di sorveglianza di massa, essa non può essere semplicemente “disinventata”. La sua esistenza stessa crea una pressione irresistibile al suo utilizzo. Se uno stato possiede armi nucleari, le altre devono averle anch’esse per difesa. Se un’azienda usa algoritmi per ottimizzare i guadagni, le concorrenti devono fare altrettanto per non soccombere. La tecnica genera il proprio imperativo d’uso, a priori delle intenzioni iniziali di chi l’ha creata.

Tale inversione della libertà si mostra anche nel rapporto dell’uomo con i dispositivi tecnici della quotidianità. Anders anticipa qui quella che diventerà la dipendenza degli smartphone e dai social media. Non è l’uomo a decidere liberamente quando e come usare la tecnica, ma è la tecnica che struttura i suoi tempi, spazi e le stesse relazioni.⁴⁷ La notifica che interrompe, l’algoritmo che suggerisce cosa guardare, tutto questo non accresce la libertà dell’uomo, ma la ridefinisce secondo parametri tecnici che sfuggono al controllo consapevole.

La creazione tecnica, quindi, non è mera creazioni di oggetti, ma la produzione di nuovi bisogni e desideri. La libertà promessa dalla tecnica (la libertà da vincoli naturali, da fatiche fisiche, limiti cognitivi), si rovescia in chiave dialettica nella sua negazione: diveniamo dipendenti da sistemi che non comprendiamo e che ci modellano secondo logiche estranee alla nostra volontà cosciente.

2.7– Il conformismo tecnico: l’uniformazione attraverso l’apparato.

Interessante del pensiero di Anders riguarda il modo in cui la tecnica produce conformismo senza bisogno di coercizione esplicita.⁴⁸ Anders osserva che nell’epoca della tecnica non è più necessario un potere centralizzato che imponga dall’alto un pensiero unico, come avveniva nei totalitarismi classici del Novecento. L’uniformazione avviene dal basso, attraverso la struttura stessa dell’apparato tecnico che media le nostre esperienze.

Prendiamo l’esempio della televisione, che Anders analizza in modo interessante. La televisione non ci dice esplicitamente cosa pensare, ma struttura come pensiamo, determinando quali eventi sono

⁴⁶ Ivi, p.298.

⁴⁷ Sul tema della "colonizzazione" del tempo di vita da parte della tecnica digitale, si veda: B. Stiegler, “La société automatique. L’avenir du travail”, Fayard, Paris 2015.

⁴⁸ G. Anders, “L’uomo è antiquato”, vol. I, cit., pp. 115-140.

degni di attenzione e quali no, quanto tempo dedicare a ciascuno, in quale ordine sequenziale fruirli.⁴⁹ Tutti gli spettatori colgono le stessi immagini, nello stesso ordine, mediate dalla medesima regia. Questo produce una “monocultura percettiva”, tutti condividiamo lo stesso mondo fantasmatico, anche se crediamo di mantenere opinioni diverse su di esso.

Tale meccanismo si è radicalizzato nell’era dei social media e degli algoritmi di raccomandazione. Mentre la tv creava un’uniformità evidente (tutti vedevano lo stesso programma), gli algoritmi creano un’uniformità mascherata da personalizzazione. Ogni utente riceve contenuti “personalizzati” in base al suo profilo, ma questa personalizzazione è fasulla, segue infatti parametri statici identici per tutti. Il risultato è un conformismo paradossale: ci sentiamo unici, ma in realtà siamo tutti sottoposti alla stessa logica di cattura dell’attenzione.⁵⁰

Anders anticipa qui quello che potremmo definire “totalitarismo morbido” della tecnica contemporanea. Non veniamo obbligati con la violenza a conformarci, ma siamo sedotti a farlo, la tecnica ci offre comodità, intrattenimento, connessione sociale, e in cambio accettiamo di delegare progressivamente le nostre scelte, gusti, e persino i nostri desideri. Non percepiamo questa delega come oppressione, ma come servizio. È proprio questa invisibilità del meccanismo di controllo lo rende pervasivo e difficile da contestare.

L’uomo oramai, afferma Anders, non è più nemmeno conscio di essere conformato perché il conformismo è incorporato nella struttura stessa degli strumenti attraverso cui accede al mondo.⁵¹

2.8 – L’apocalisse senza apocalisse: la catastrofe come mutamento ontologico.

Una delle tesi più profonde del pensiero di Anders riguarda la sua concezione di apocalisse nell’epoca della tecnica. Anders ribalta completamente la comprensione originale del termine “apocalisse”, che nella storia del pensiero occidentale, (dalle narrazioni bibliche alle visioni profetiche del Medioevo, fino alle fantasie fantascientifiche del Novecento), è sempre stata immaginata come un evento drammatico, che segna nettamente un prima da un dopo.⁵²

Per Anders invece l’apocalisse della tecnica non è un evento futuro da temere, ma una condizione già presente, un cambiamento ontologico della struttura stessa dell’esistenza umana che è già avvenuto

⁴⁹ Ivi, pp. 125-128. Anders parla della televisione come di un "mondo pre-masticato"

⁵⁰ Su questo tema si veda: S. Zuboff, “Il capitalismo della sorveglianza”, Luiss University Press, Roma 2019 (*ed. or.* The Age of Surveillance Capitalism, 2019)

⁵¹ Cfr. H. Marcuse, “L’uomo a una dimensione”, Einaudi, Torino 1967, pp. 25-30. Marcuse, contemporaneo di Anders, sviluppa temi convergenti sul "totalitarismo tecnologico" delle società industriali avanzate

⁵² G. Anders, “L’uomo è antiquato”, vol. II, cit., pp. 233-270. Il capitolo "La cecità verso l’apocalisse"

senza che ce ne accorgessimo. La formula con cui esprime questo concetto è ‘‘Apokalypse ohne Apokalypse’’ (apocalisse senza apocalisse)⁵³. Non c’è stato un momento preciso, una data identificabile in cui tutto è cambiato. Eppure, tutto è cambiato. L’uomo ha perduto qualcosa di essenziale, il controllo sui processi che ha messo in moto, la capacità di rappresentarsi le conseguenze delle proprie azioni, la possibilità stessa di essere un soggetto responsabile. Ma questa perdita non è avvenuta con un evento preciso, ma attraverso una trasformazione graduale e impercettibile.

Il caso concreto che Anders usa per mostrare questo concetto è quello della bomba atomica. L’apocalisse nucleare, nella comprensione comune, è l’evento futuro temuto: l’esplosione che distrugge tutto. Ma Anders ribalta la prospettiva. L’apocalisse è già avvenuta nell’istante stesso in cui la bomba è stata creata, nel momento in cui è diventato tecnicamente auspicabile annientare l’intera specie umana. Da quell’istante preciso, agosto 1945, Hiroshima e Nagasaki, l’intera umanità vive in una condizione totalmente diversa da quella di tutte le generazioni precedenti. Per la prima volta nella storia, l’umanità ha i mezzi tecnici della propria auto-distruzione. Tale possibilità, una volta realizzata, non può più essere eliminata. La conoscenza di come costruire armi nucleari è irreversibile; anche se tutte le bombe esistenti venissero distrutte, il sapere tecnico rimarrebbe, pronto ad essere riattivato in qualsiasi momento.

Ciò che si è trasformata, quindi, è lo statuto ontologico dell’umanità. Non c’è più un futuro garantito, non c’è più la sicurezza che le generazioni prossime esisteranno. Questa trasformazione della condizione umana è apocalittica nel senso più profondo del termine. È una trasformazione che è avvenuta silenziosamente, attraverso un progetto tecnico e scientifico che veniva celebrato come trionfo dell’intelligenza umana.

Anders riflette su questa analisi dimostrando come l’apocalisse nucleare sia strutturalmente irrepresentabile per la coscienza umana. Non possiamo immaginare concretamente, la nostra estinzione, non possiamo farcene una un’immagine vissuta.⁵⁴ Questa sproporzione tra la portata oggettiva dell’evento possibile e la nostra capacità soggettiva di rappresentarcelo è, di nuovo, il dislivello prometeico applicato alla sua forma più estrema.

Il risultato è quello che Anders chiama: cecità verso l’apocalisse. L’essere umano non riesce a cogliere la catastrofe perché non ha la forma attesa. Si aspetta un evento, un momento di rottura, e quindi non riconosce la trasformazione ontologica già successa. L’uomo continua ad esistere come se la minaccia nucleare fosse un problema tra gli altri, uno dei pericoli da gestire, quando in realtà segna la fine di un’epoca della storia umana e l’ingresso di una condizione totalmente nuova. Si vive oggi nell’era

⁵³ Ivi, p. 250

⁵⁴ G. Anders, ‘‘Essere o non essere. Diario di Hiroshima e Nagasaki’’, Einaudi, Torino 1961, pp. 15-25

post-apocalittica senza saperlo, poiché l'immaginazione umana resta ferma ai modelli tradizionali dell'apocalisse come evento futuro.

Quindi, nell'epoca pre-tecnica il futuro era aperto e garantito nella sua struttura fondamentale, nell'epoca della tecnica matura, questa certezza è rovesciata; il futuro è incerto nella sua stessa esistenza. Non sappiamo se ci sarà un futuro⁵⁵. Questo ribaltamento della temporalità ha effetti esistenziali profonde. Se non possiamo più proiettarci in un futuro certo, allora la struttura stessa della progettualità umana (quella che Heidegger chiamava il carattere progettuale dell'esistenza); viene minata alla base.

Anders parla di una "espropriazione del futuro". La tecnica ci ha tolto non solo il controllo sul presente, ma anche quella di abitare il futuro. Ogni generazione vissuta prima poteva immaginare che, qualsiasi cosa fosse accaduta, l'umanità sarebbe continuata. Questa iscrizione nella continuità temporale della specie era una delle fonti essenziali di senso dell'esistenza umana. Oggi questa fonte è prosciugata.

Un altro aspetto essenziale della concezione di Anders, è che l'apocalisse non si manifesta con un unico evento (l'esplosione nucleare), ma è un processo che opera su più livelli contemporaneamente. La bomba atomica è solamente la manifestazione più evidente e drammatica di una trasformazione più profonda. Anders identifica diverse forme di apocalisse umana, che procedono simultaneamente, ognuna delle quali contribuisce alla dissoluzione progressiva di ciò che definisce l'esistenza umana come l'abbiamo conosciuta.⁵⁶

Teorizza l'apocalisse fisica (la minaccia di distruzione materiale con le armi di annientamento di massa), l'apocalisse percettiva (in cui si costituisce il mondo dei fantasmi mediativi, l'atrofia della capacità di esperienza diretta). C'è poi l'apocalisse morale, in cui si va a frammentare la responsabilità, l'azione come funzione dentro ad un apparato, il cosiddetto nichilismo tecnico, ed infine l'apocalisse ontologica (la riduzione dell'uomo ad essere cosa, perdita di unicità). Tutte queste dimensioni di apocalisse procedono simultaneamente, si rafforzano una con l'altra, e nessuna di esse assume la forma di evento singolare. Sono trasformazioni che avvengono nel tempo in modo continuativo e graduale, andando a ridefinire il concetto di essere umano.

Di fronte a tale situazione, Anders propone quello che potremmo definire un compito filosofico nell'era della tecnica: creare una "immaginazione apocalittica"⁵⁷. Questa immaginazione richiede di compiere uno sforzo attivo di rappresentazione, cercare di cogliere la catastrofe mentre succede, di riconoscerla nella normalità della quotidianità, di cogliere la perdita che si maschera come guadagno.

⁵⁵ Anders II, pp. 260-265.

⁵⁶ Anders sviluppa questa tesi del "pluri-apocalisse" in diversi saggi sparsi, raccolti poi in parte in *Mensch ohne Welt* (1984)

⁵⁷ Anders II, pp. 267-270

È un esercizio contro-intuitivo che richiede di resistere alla pressione del senso comune che identifica automaticamente progresso tecnico e progresso per la condizione umana.

L'immaginazione apocalittica non vuol dire fantasticare scenari pessimistici, ma compiere una sorta di de-familiarizzazione del presente. È un'operazione filosofica che richiede di sospendere l'abitudine, di rifiutare la normalizzazione.

Anders, d'altra parte, sa quanto la potenza della tecnica mina la possibilità di mantenere attiva la coscienza critica. Eppure, sottolinea Anders, è l'unico modo adeguato alla situazione. Se non riusciamo a cogliere l'apocalisse che è già in corso, non avremo nessuna possibilità di opporvi resistenza. La cecità verso l'apocalisse assicura il suo compimento.

Capitolo 3- La deriva della tecnica nel presente: Turing e Anders e la questione.

3.1 – Il malinteso come sintomo: bug tecnico o abisso ontologico?

Uno dei modi più efficaci per capire la differenza tra le posizioni di Turing e Anders, è riflettere come i due interpreterebbero gli errori e i fraintendimenti che emergono continuamente nel rapporto uomo e sistemi di IA. È fondamentale chiarire cosa si intende con questo termine: i sistemi di intelligenza artificiale attuali, in particolare quelli capace di creare testo o rispondere a domande, commettono errori molto particolari che non assomigliano ai classici errori di calcolo di un computer. Creano citazioni inesistenti presentandole come reali, attribuiscono ad autori opere che non hanno mai composto, descrivono eventi mai accaduti con una precisione che rende difficoltoso dubitarne.⁵⁸ Questi non sono casi rari o eccezionali, ma si tratta di un comportamento tipico, che emerge ogni volta che il sistema non dispone di abbastanza informazioni ma è strutturato per dare comunque una risposta. Oltre a questo tipo di errore, esistono incomprensioni di natura diversa. Il sistema interpreta un quesito ironico come letterale, non coglie il tono di una frase, non riconosce metafore o riferimenti culturale che qualsiasi parlante umano coglierebbe subito. E poi c'è un terzo tipo di problema, che può essere considerato il più interessante filosoficamente, il sistema può descrivere con precisione l'esperienza di tenere in mano un oggetto pensante, di sentire freddo, di essere stanco, senza aver provato mai queste esperienze. Gestisce simboli linguistici che rimandano ad esperienze corporee, senza mai aver avuto un corpo. Il quesito filosoficamente cruciale non è descrivere questi errori, ma comprendere cosa ci dicono. Sono problemi tecnici volti ad essere risolti, oppure sono il sintomo di qualcosa che va più a fondo?

La risposta di Turing sarebbe senza dubbio la prima. La sua visione, tali errori sono imperfezioni, non limiti insuperabili. Se il sistema inventa citazioni inesistenti, è perché i dati su cui è stato addestrato erano insufficienti o sbilanciati, la soluzione è più dati, di qualità superiore. Se non comprende l'ironia, è perché l'architettura attuale non cattura certe strutture del linguaggio: la soluzione è un'architettura più sofisticata. Se non ha accesso all'esperienza corporea, è perché fino ad ora i sistemi sono stati quasi esclusivamente linguistici, e quindi la soluzione è integrare visione, udito, e capacità di interazione fisica con il mondo.⁵⁹ In questa visione il malinteso è rumore nel sistema, un'imperfezione temporanea. Turing lo aveva già detto confutando l'obiezione di Lady Lovelace, in cui aveva affermato che i limiti delle macchine attuali non ci dicono nulla sui limiti di principio delle macchine future.

⁵⁸ Cfr. Z. Ji et al., "Survey of Hallucination in Natural Language Generation", in «ACM Computing Surveys», 55, 12, 2023, pp. 1-38.

⁵⁹ A. Turing, "Computing Machinery and Intelligence", in «Mind», LIX, 236, 1950, p. 456.

Anders vedrebbe le cose in modo opposto. Questi errori non sono semplice bug da correggere, ma sono la manifestazione pragmatica del dislivello prometeico, la prova che stiamo generando sistemi che operano in un universo di significati strutturalmente diverso dal nostro. Il nucleo centrale è la distinzione, già analizzata nel capitolo precedente, tra il mondano e il fantasma. Mondano è la realtà che conosciamo attraverso il corpo, la resistenza fisica degli oggetti, il peso delle cose, il caldo e il freddo, il dolore⁶⁰. Il fantasma è invece l'immagine tecnica, la rappresentazione mediata, il dato digitale che sostituisce l'esperienza diretta. Il sistema che crea testo opera interamente nel dominio dei fantasmi: elabora pattern statistici estratti da milioni di testi, senza aver mai avuto accesso al mondo che quei testi descrivono. Per Anders questo abisso non è colmabile con più dati o con architetture più sofisticate, perché non è una questione quantitativa ma qualitativa. Fa riferimento al tipo di rapporto con il reale, non la qualità di informazioni disponibili.

C'è però un aspetto di questi errori che va aldilà della questione della comprensione e tocca direttamente il problema della responsabilità. Quando un sistema produce un errore che danneggia qualcuno, che dà informazioni false che qualcuno usa per prendere una decisione sbagliata, chi risponde? Il programmatore che ha scritto il codice anni prima? O l'azienda? Questa frammentazione della catena causale e morale è esattamente il meccanismo che Anders aveva descritto analizzando Eichmann. Nessuno si sente responsabile perché ognuno ha semplicemente portato a termine il compito che gli era stato assegnato, in un apparato più grande.⁶¹

Le due interpretazioni del malinteso mostrano due filosofie opposte. Per Turing il malinteso è un problema ingegneristico sulla via verso sistemi sempre migliori. Per Anders è la prova che abbiamo creato qualcosa che non controlliamo e non comprendiamo totalmente, e il tentativo di risolvere il problema con più tecnica non fa che approfondire il dislivello anziché colmarlo.

3.2 – Dalla bomba atomica all'intelligenza artificiale: la logica dell'incontrollabilità.

Il parallelismo tra bomba atomica e intelligenza artificiale non va inteso come analogia retorica tra due tecnologie distruttive, ma come confronto tra due forme di potere tecnico che vanno altro la capacità ordinario di previsione e controllo. Ciò che le accomuna non è l'effetto in quanto tale, ma la struttura della possibilità che inaugurano.

La bomba atomica ha introdotto per la prima volta nell'umanità la possibilità concreta dell'annientamento globale. Come è stato colto nel dibattito successivo alla Seconda guerra mondiale,

⁶⁰ G. Anders, "L'uomo è antiquato", vol. I: Sull'anima nell'epoca della seconda rivoluzione industriale, Bollati Boringhieri, Torino 2002 (ed. or. 1956), pp. 115-140.

⁶¹ G. Anders, "Noi figli di Eichmann", Giuntina, Firenze 1995 (ed. or. 1964), pp. 34-55

essa non rappresenta solamente un'arma più forte, ma una trasformazione del rapporto tra potenza tecnica e sopravvivenza collettiva⁶². Il suo effetto non si esaurisce nell'esplosione; esso si estende alle conseguenze indirette e sistematiche, che sfuggono alla volontà del singolo attore.

L'intelligenza artificiale non produce distruzione fisica immediata, ma introduce una diversa forma di pervasività. La sua capacità di incidere contemporaneamente su ambiti economici, informativi e decisionali li rende una struttura infrastrutturale. A differenza della bomba, che agisce in modo preciso e catastrofico, l'IA opera in modo diffuso e continuo. Il suo potere non è solo nell'evento, ma distribuito nei processi.

Il punto d'incontro tra due tecnologie sta nella difficoltà di anticiparne totalmente le conseguenze sistematiche. Nel campo nucleare, la scoperta della possibilità di un "inverno nucleare" ha mostrato come gli effetti di un conflitto atomico potessero superare di molto le intenzioni iniziali.⁶³ Similmente, nel campo dell'IA, molti studiosi hanno sottolineato come sistemi progettati per ottimizzare determinati fini possano generare esiti non previsti o indesiderati quando operano su larga scala.⁶⁴ Questa dinamica è stata descritta, in ambito nozionistico, attraverso il concetto di emergenza: proprietà che non sono direttamente deducibili dai singoli componenti, ma che si mostrano nell'interazione complessa del sistema.⁶⁵ Nel caso dell'IA oggi, l'emergenza si traduce in comportamenti che vanno oltre la programmazione esplicita, rendendo più complicato stabilire un nesso lineare tra intenzione progettuale ed effetto reale.

La questione dell'incontrollabilità, quindi, non fa riferimento soltanto alla potenza, ma non-linearità. Quanto più un sistema diviene complesso e integrato in infrastrutture sociali, tanto più i suoi effetti si distribuiscono in modo imprevedibile. Questo vale per i mercati finanziari automatizzati, per i sistemi di raccomandazione che modellano l'accesso all'informazione, per gli algoritmi di selezione utilizzati in ambito lavorativo o giudiziario.

Un altro elemento di analogia è rappresentato dalla difficoltà immaginativa di rappresentare totalmente gli scenari più estremi. Nel dibattito degli ultimi anni sull'intelligenza artificiale (AGI), alcuni autori hanno richiamato l'attenzione sul problema dell'"alignment", ossia sull'allineamento tra fini del sistema e valori dell'essere umano⁶⁷. La paura non è riferita solamente sull'uso malevolo della tecnologia, ma la possibilità un sistema molto autonomo persegua in modo coerente un fine formalizzato senza dividerne il contesto etico implicito.

⁶² H. Jonas, "Il principio responsabilità", Torino, Einaudi, 1990

⁶³ Cfr. C. Sagan et al., "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions", *Science*, 222 (1983), pp. 1283–1292.

⁶⁴ N. Bostrom, "Superintelligence: Paths, Dangers, Strategies," Oxford University Press, 2014

⁶⁵ M. Mitchell, "Complexity: A Guided Tour," Oxford University Press, 2009.

⁶⁷ S. Russell, "Human Compatible: Artificial Intelligence and the Problem of Control", Viking, 2019.

Questa paura non implica l'imminenza di scenari apocalittici, ma segna un limite di struttura: la difficoltà di garantire che un sistema operante su scala globale rimanga sempre riconducibile ad un controllo umano significativo. L'incontrollabilità non coincide con il caos, può manifestarsi come rigidità algoritmica, come automatizzazione di criteri parziali, come amplificazione di bias preesistenti.⁶⁸

La differenza tra bomba e IA resta evidente sul piano materiale, ma l'analogia si potenzia sul piano simbolico e politico. Tutte e due rappresentano il punto in cui la capacità tecnica supera la capacità ordinaria di rappresentazione degli effetti. L'effetto più significativo non è soltanto ciò che fanno, ma ciò che rendono possibile fare.

In tale senso, la domanda non è l'IA distruggerà il mondo, ma se la sua integrazione sistemica possa produrre trasformazioni profonde nelle condizioni di autonomia, lavoro, informazione.

L'incontrollabile, quindi, assume una forma diversa rispetto all'evento nucleare. Non è l'esplosione, ma l'erosione graduale di margini decisionali. Non è collasso immediato ma la sua trasformazione silenziosa nella realtà sociale. Proprio per questo, la sua visione risulta più ambigua. Mentre la minaccia atomica era visibile nella sua drammaticità, quella algoritmica si presenta come innovazione.

Il confronto tra le due tecnologie permette così di isolare una struttura comune: la scissione tra potenza tecnica e capacità di governo etica. Quando la prima cresce più velocemente della seconda, si crea uno scarto che può tradursi in vulnerabilità sistematica. L'incontrollabilità non è necessariamente perdita totale di controllo, ma riduzione della capacità di prevedere e orientare fini complessivi.

In questa visione, il problema non è fermare lo sviluppo tecnologico, ma interrogare le condizioni del suo governo. La domanda fondamentale diventa politica e normativa: quali limiti, quali meccanismi di supervisione, forme di responsabilità collettiva possono garantire che la crescita della potenza algoritmica non si traduca in una delega irreversibile della decisione?

3.3 – Il caso D-Syde: esperienza tecnica e coscienza del limite.

Per verificare concretamente la tenuta concreta delle ipotesi teoriche discusse nei capitoli precedenti e osservare se le categorie filosofiche elaborate da Turing e Anders abbiano una qualche risonanza nell'esperienza vissuta, è opportuno integrare l'analisi puramente teorica con un'indagine empirica di carattere qualitativo condotta direttamente sul campo. L'obiettivo specifico è stato quello di osservare come le categorie astratte di ottimismo del progetto, e di pessimismo delle conseguenze, si manifestino concretamente nella prassi quotidiana di chi opera professionalmente nel cuore pulsante

⁶⁸ B. D. Mittelstadt et al., "The Ethics of Algorithms: Mapping the Debate", *Big Data & Society*, 2016, vol.3, n.2.

dell'innovazione tecnologica contemporanea, costruendo e mantenendo sistemi di intelligenza artificiale.

L'indagine è stata condotta presso D-Syde, un'azienda di sviluppo software con sede a Villorba, in provincia di Treviso. Il campione, sebbene circoscritto a sette professionisti del settore, risulta particolarmente significativo per la profondità e la varietà delle competenze tecniche coinvolte. Si tratta di figure professionali attivamente impegnate nella progettazione, sviluppo e gestione di sistemi informatici complessi. Il questionario è stato somministrato in forma anonima con Google Moduli, proprio per garantire la massima sincerità possibile nelle risposte, specialmente su questioni eticamente e professionalmente sensibili come quella della responsabilità personale davanti agli errori dei sistemi sviluppati.

Il primo nodo cruciale di indagine ha riguardato direttamente la questione ontologica essenziale sulla natura della comprensione nelle macchine. Alla domanda esplicita sulla reale capacità di comprensione dei sistemi di intelligenza artificiale su cui lavorano quotidianamente, se questi sistemi comprendano davvero ciò che elaborano oppure si limitino a seguire pattern statistici complessi senza cogliere il significato autentico, la maggioranza netta dei partecipanti, precisamente il 60%, ha risposto con estremo pragmatismo indicando che i sistemi non comprendono nel senso comune e pieno del termine, ma operano attraverso l'individuazione e la manipolazione di pattern statistici complessi estratti dai dati. Solo il 20% ha attribuito ai sistemi una qualche forma di comprensione, mentre un altro 20% ha scelto una posizione intermedia indicando che si tratta di un mix di entrambe le cose.

Questo dato empirico è fondamentale importanza filosofica: anche tra chi lavora tutti i giorni in quel campo tecnico specifico, tra chi ha una conoscenza intima e pratica del funzionamento pragmatico di questi sistemi, emerge chiaramente la consapevolezza critica di una fondamentale mancanza di competenza referenziale autentica da parte della macchina. Nonostante l'efficacia operativa impressionante del software, che può produrre testi coerenti e informativi, continua a persistere tra i professionisti la percezione netta di un abisso qualitativo insuperabile tra la manipolazione puramente sintattica dei segni linguistici e l'accesso semantico genuino al mondo concreto che quei segni dovrebbero rappresentare. La macchina, quindi, viene colta e compresa dai suoi stessi creatori come uno strumento di potenza fondamentalmente privo di quella dimensione di intenzionalità, di riferimento al mondo, di comprensione autentica del significato che contraddistingue costitutivamente l'essere umano pensante e parlante.

Un altro aspetto importante che emerge dall'indagine riguarda l'esperienza diretta dell'imprevedibilità e dell'emergenza di comportamenti non anticipati. Una parte molto consistente dei professionisti intervistati ha ammesso esplicitamente di aver incontrato nella propria pratica

lavorativi comportamenti che possono essere definiti emergenti, cioè risposte o pattern di funzionamento completamente inattesi e non previsti esplicitamente nell'intenzione iniziale. L'imprevisto, in questo contesto tecnico e specifico, non viene più percepito come un malfunzionamento o un errore da correggere, ma viene riconosciuto come una proprietà intrinseca e non eliminabile dalla complessità stessa degli algoritmi di apprendimento automatico contemporanei. Questa ammissione sincera e significativa apre la strada ad una riflessione filosofica molto profonda sul concetto stesso di autonomia tecnica. Lo sviluppatore, non è più il dominatore assoluto del codice che prodotto, ma si trasforma progressivamente e inevitabilmente in una sorta di supervisore esterno di una serie di processi in cui gli effetti non sono sempre prevedibili. Si assiste quindi, anche nell'esperienza soggettiva vissuta dai professionisti stessi, ad un cambiamento qualitativo della relazione tra l'uomo e la macchina, in cui la tecnica sembra acquisire una propria logica interna di sviluppo e adattamento che procede in modo relativamente autonomo rispetto alla volontà progettuale consapevole e al controllo diretto.

L'indagine ha toccato un tema molto profondo in chiave etica; la questione della responsabilità morale e legale. Di fronte allo scenario ipotetico ma molto realistico di un errore del sistema con effetti negativi concreti per qualcuno, le risposte hanno sottolineato in modo drammatico una profonda frammentazione dei ruoli e un'eliminazione del senso di responsabilità individuale. La difficoltà enorme, e forse l'impossibilità pratica, di individuare un responsabile unico e chiaramente identificabile, scegliendo tra programmatore individuale, azienda, team di raccolta dati, e utilizzatore finale, conferma pienamente la tendenza alla parcellizzazione dell'azione tecnica che Anders aveva diagnosticato come caratteristica strutturale dell'epoca della tecnica matura.

Questa condizione esistenziale e professionale ricorda l'analisi che Anders aveva sviluppato sulla figura storica di Eichmann: l'uomo individuale esegue una funzione parziale circoscritta all'interno di un apparato tecnico molto più grande e complesso, senza mai percepirsi veramente come soggetto morale totalmente responsabile del risultato complessivo finale prodotto dall'apparato nella sua totalità. La frase ricorrente "non è colpa mia, ho solo fatto il mio lavoro" trova qui una tradizione contemporanea perfetta nel dominio dell'IA: il programmatore ha scritto il codice, ma non ha scelto personalmente i dati di addestramento. Chi ha raccolto i dati non ha progettato l'architettura dell'algoritmo; chi ha progettato l'architettura non decide come viene usato il sistema di produzione. L'algoritmo stesso decide secondo logiche interne opache che nessuno comprende totalmente. L'effetto è una dissoluzione quasi completa della responsabilità morale che nessuna riforma sembra in grado di ricostruire pienamente.

Per quanto riguarda la discussione di questioni etiche nei luoghi di lavoro, i dati sono allarmanti. La maggioranza degli intervistati ha dichiarato che le questioni etiche e le conseguenze sociali dei sistemi

su cui lavorano vengono discusse raramente o mai. Questo conferma quanto Anders aveva colto: l'apparato tecnico produce negli individui una concentrazione esclusiva sulla propria funzione parziale, rendendo difficile la riflessione d'insieme necessaria per il giudizio morale.

Non è che i programmatori siano privi di sensibilità etica. Il punto è più sottile, i contesti lavorativi in cui operano non prevedono spazi e tempi per questa riflessione. La pressione verso la produttività, la competizione nel settore, la logica dell'apparato che richiede efficienza fanno sì che la domanda etica sul senso di ciò che si sta costruendo venga sistematicamente accantonata, non per scelta consapevole ma per inerzia del sistema stesso.

Sulla scala di ottimismo e pessimismo riguardo al futuro dell'intelligenza artificiale, che andava da 1 (massimo pessimismo) a 10 (massimo ottimismo), la distribuzione delle risposte è significativa. La maggior parte si è collocata su valori medio alti, tra 7 e 9 indicando un ottimismo moderato. Ma c'è stata anche una risposta che ha segnato 1, esprimendo un pessimismo totale. Questa distribuzione mostra che anche all'interno dello stesso piccolo gruppo di professionisti coesistono visioni del mondo agli estremi opposti del confronto tra l'ottimismo di Turing e il pessimismo di Anders.

Nel complesso, l'indagine conferma che la tensione tra Turing e Anders non è una disputa astratta tra posizioni storiche superate. È una tensione che si riproduce nell'esperienza quotidiana di chi costruisce questi sistemi. La maggioranza sperimenta fenomeni che le categorie di Anders descrivono con precisione: perdita di controllo, assenza di riflessione etica, frammentazione della responsabilità, vertigine di fronte alla potenza di ciò che si crea. Al tempo stesso, l'ottimismo tecnico resta presente, suggerendo che le due prospettive coesistono, spesso in modo contraddittorio, anche nella coscienza delle stesse persone.

Conclusion

Il percorso di questa tesi è iniziato da una domanda che nel 1950, quando Alan Turing la formulò, sembrava ancora speculativa; ‘‘le macchine possono pensare?’’ settantacinque anni dopo, questo quesito ha ricevuto una risposta largamente affermativa sul piano tecnico. I sistemi di IA parlano in linguaggio naturale, risolvono problemi complessi, apprendono dall’esperienza, creano contenuti originali. Molte delle previsioni di Turing si sono realizzate con precisione impressionante. Sul piano delle capacità operative, il suo ottimismo ha trovato conferma nella realtà tecnologica contemporanea. Eppure, come questa tesi ha cercato di cogliere attraverso il confronto sistematico con Gunther Anders, rispondere affermativamente alla domanda di Turing non chiude il problema filosofico, ma lo apre in modo ancora più urgente. Affianco alla questione tecnica sulla possibilità del pensiero artificiale, nasce una domanda diversa e più profonda: cosa accade a noi, uomini, quando le macchine pensano? Questa seconda domanda fa riferimento alle conseguenze esistenziali, morali, politiche della tecnica per la condizione umana. Ed è qui che l’ottimismo del progetto si scontra inevitabilmente con il pessimismo delle conseguenze, che la deriva della tecnica diventa visibile come fenomeno reale del nostro presente.

La deriva non è da intendere come catastrofe futura da evitare ma una struttura già operante. È lo scarto tra ciò che progettiamo consapevolmente e ciò che produciamo effettivamente, tra le intenzioni esplicitate dei creatori e gli effetti concreti su chi ne subisce gli effetti, tra il successo tecnico misurabile e le trasformazioni antropologiche che quel successo porta con sé. Turing progettava macchine intelligenti per liberare l’essere umano dai compiti ‘‘meccanici’’, per accrescere le sue capacità e per produrre il progresso. Quest’ultimo si è realizzato sul piano tecnico. Ma sono nati effetti collaterali che nulla sembra in grado di eliminare: la frammentazione della responsabilità in sistemi dove nessuno si sente veramente responsabile del risultato finale, la perdita progressiva di controllo su processi che i loro stessi creatori descrivono come spettatori più che come progettisti, la riproduzione a livello digitale di quel meccanismo che Anders aveva identificato analizzando la figura di Eichmann.

Il confronto tra le due posizioni ha mostrato che la loro differenza non riguarda principalmente i fatti tecnici. Tutti e due riconoscono che le macchine cambiano profondamente l’essere umano e il suo modo di esistere. Divergono nel giudizio di valore su questa trasformazione. Per Turing è liberatoria. Un superamento dei limiti naturali dell’uomo verso forme più avanzate di intelligenza. Per Anders è alienante, una perdita di autonomia, di giudizio, di quella capacità di essere soggetto morale responsabile che costituisce il nucleo dell’umano. Questa divergenza non è risolvibile con nuovi dati

o più ricerca tecnica, perché non è una disputa scientifica ma un conflitto tra visioni della realtà, tra concezioni diverse di cosa valga la pena preservare nella condizione umana.⁶⁹

Il caso studio presso D-Syde ha dato conferma empirica che queste tensioni teoriche ci siano nell'esperienza quotidiana di chi costruisce questi sistemi. La maggioranza dei professionisti intervistati sperimenta fenomeni che Anders descrive con precisione: percezione di perdita di controllo sul processo tecnico, assenza di spazi per la riflessione etica nei contesti lavorativi, frammentazione della responsabilità che rende impossibile identificare un responsabile unico, vertigine davanti alla potenza di quello che si costruisce senza comprenderlo del tutto. Al tempo stesso, l'ottimismo tecnico resta presente, suggerendo che le due visioni non si escludono ma coesistono, molte volte in modo contraddittorio, anche nella coscienza delle stesse persone.

Questa coesistenza contraddittoria non è un difetto di coerenza logica, ma il riflesso fedele di una ambiguità costitutiva della tecnica contemporanea. La stessa realtà tecnologica può essere colta legittimamente sia come promessa di progresso sia come minaccia esistenziale. Turing e Anders non sono teoria da valutare come vere o false, ma mostrano due aspetti reali della nostra condizione. Turing aveva ragione sul piano delle possibilità in ambito tecnologico. Anders aveva ragione sul piano degli effetti antropologici. E il nucleo centrale è che il successo tecnico non toglie il problema antropologico, anzi lo rende ancora più urgente. Più i sistemi diventano forti ed efficaci, più diventa complicato comprenderli, controllarli e assumersi la responsabilità delle loro conseguenze.

La questione con cui la tesi si chiude è quella che Anders ci ha insegnato a porci: possiamo permetterci che le macchine pensino? Siamo in grado, di abitare consapevolmente il mondo tecnico che stiamo costruendo a velocità accelerata? Abbiamo sviluppato le istituzioni politiche, i linguaggi morali necessarie per governare gli effetti di questa capacità tecnica che continuiamo a potenziare?⁷⁰

La deriva della tecnica non è un destino ineluttabile a cui rassegnarsi in modo passivo. Ma nemmeno si risolverà automaticamente con progresso tecnologico. Richiede una presa di coscienza collettiva della posta in gioco, un esercizio difficile di immaginazione morale. Una volontà politica di porre limiti e orientare in modo consapevole lo sviluppo invece di subirlo come necessità tecnica inevitabile. Richiede di invertire il rapporto abituale: non domandarci cosa la tecnica può fare e poi adattarci agli effetti, ma domandarci quale mondo vogliamo abitare e poi selezionare gli sviluppi tecnici compatibili con questa visione.

Questa inversione, però, non è solamente una questione di buona volontà individuale o di scelte di consumo personali. Richiede trasformazioni strutturali nelle condizioni che creano la tecnica. Spazi

⁶⁹ Sul conflitto tra valori nell'interpretazione della tecnica cfr. L. Winner, *The Whale and the Reactor: A Search for Limits in an "Age of High Technology,"* University of Chicago Press, Chicago 1986, pp. 3-18.

⁷⁰ G. Anders, *L'"uomo è antiquato,"* vol. II, tr. it. M.A. Mori, Bollati Boringhieri, Torino 2003, pp. 295-310.

pubblici di discussione in cui le scelte tecnologiche possano essere dibattute democraticamente, formazione che non sia solo tecnica ma anche filosofica ed etica. Richiede di riconoscere che le questioni tecnologiche sono sempre anche questioni politiche, e che delegare agli esperti o al mercato vuol dire aver già fatto una scelta politica molto precisa.

Se questa trasformazione sia ancora possibile, e come, resta la domanda aperta più urgente. Ma ciò che è sicuro è che senza il tentativo di recuperare la capacità di sentire moralmente quello che produciamo tecnicamente, senza il coraggio di porre limiti allo sviluppo quando necessario, senza la volontà di subordinare la logica tecnica ed economica a quella etica e politica, il destino prospettato da Anders diventa sempre più probabile. L'essere umano diventa obsoleto non perché sostituito dalle macchine, ma perché ha rinunciato volontariamente a quella dimensione di autonomia, responsabilità e giudizio critico che lo rendeva propriamente umano.

La deriva della tecnica, quindi, non è solo un fenomeno da analizzare filosoficamente, ma una sfida esistenziale e politica da affrontare collettivamente. Non si richiede di scegliere una posizione tra quella di Turing e Anders, ma di riconoscere che tutte e due descrivono aspetti reali del nostro presente che il compito urgente è costruire le condizioni per governare consapevolmente questa tensione invece di subirla. Questo è il contributo che la filosofia può ancora dare nell'epoca della tecnica: non fornire risposte definitive, ma mantenere aperte le domande essenziali, quelle che la pressione verso l'efficienza e la produttività tende sistematicamente ad eliminare. E la questione fondamentale rimane: che tipo di esseri umani vogliamo essere, e quale mondo vogliamo abitare?

Bibliografia

- Anders, Günther, "Essere o non essere. Diario di Hiroshima e Nagasaki", tr. it. E. Picco, Einaudi, Torino 1961 (ed. or. "Off Limits für das Gewissen", Rowohlt, Hamburg 1961).
- Anders, Günther, "L'uomo è antiquato". Vol. I: "Sull'anima nell'epoca della seconda rivoluzione industriale", tr. it. L. Dallapiccola, Bollati Boringhieri, Torino 2002 (ed. or. "Die Antiquiertheit des Menschen", vol. I, C.H. Beck, München 1956).
- Anders, Günther, "L'uomo è antiquato". Vol. II: "Sulla distruzione della vita nell'epoca della terza rivoluzione industriale", tr. it. M. A. Mori, Bollati Boringhieri, Torino 2003 (ed. or. "Die Antiquiertheit des Menschen", vol. II, C.H. Beck, München 1980).
- Anders, Günther, "Mensch ohne Welt. Schriften zur Kunst und Literatur", C.H. Beck, München 1984.
- Anders, Günther, "Noi figli di Eichmann", tr. it. A. Guccione Monroy, Giuntina, Firenze 1995 (ed. or. "Wir Eichmannsöhne. Offener Brief an Klaus Eichmann", C.H. Beck, München 1964).
- Bauman, Zygmunt, "Modernità e Olocausto", tr. it. M. Baldini, Il Mulino, Bologna 1992 (ed. or. "Modernity and the Holocaust", Polity Press, Cambridge 1989).
- Bender, Emily M., Gebu, Timnit, McMillan-Major, Angelina, Shmitchell, Shmargaret, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?", "Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT)", ACM, New York 2021, pp. 610–623.
- Block, Ned, "What is Functionalism?", in N. Block (a cura di), "Readings in Philosophy of Psychology", vol. 1, Harvard University Press, Cambridge (MA) 1980, pp. 171–184.
- Bostrom, Nick, "Superintelligence: Paths, Dangers, Strategies", Oxford University Press, Oxford 2014.
- Copeland, Brian Jack, "The Essential Turing", Oxford University Press, Oxford 2004.
- Dries, Christian, "Günther Anders's Philosophical Anthropology: Existential Threat and the Question of Praxis", "Thesis Eleven", vol. 120, n. 1, 2014, pp. 27–47.
- Ji, Ziwei, Lee, Nayeon, Frieske, Rita, Yu, Tiezheng, Su, Dan, Xu, Yan, Ishii, Etsuko, Bang, Yejin, Madotto, Andrea, Fung, Pascale, "Survey of Hallucination in Natural Language Generation", "ACM Computing Surveys", vol. 55, n. 12, 2023, pp. 1–38.
- Jonas, Hans, "Il principio responsabilità. Un'etica per la civiltà tecnologica", tr. it. P. Rinaudo, Einaudi, Torino 1990 (ed. or. "Das Prinzip Verantwortung", Insel Verlag, Frankfurt am Main 1979).

- Koch, Christof, Massimini, Marcello, Boly, Mélanie, Tononi, Giulio, "Neural correlates of consciousness: progress and problems", "Nature Reviews Neuroscience", vol. 17, 2016, pp. 307–321.
- LeCun, Yann, Bengio, Yoshua, Hinton, Geoffrey, "Deep Learning", "Nature", vol. 521, 2015, pp. 436–444.
- Marcuse, Herbert, "L'uomo a una dimensione", tr. it. L. Gallino e T. Giani Gallino, Einaudi, Torino 1967 (ed. or. "One-Dimensional Man", Beacon Press, Boston 1964).
- Minsky, Marvin, "Steps Toward Artificial Intelligence", "Proceedings of the IRE", vol. 49, n. 1, 1961, pp. 8–30.
- Mitchell, Melanie, "Complexity: A Guided Tour", Oxford University Press, Oxford 2009.
- Mittelstadt, Brent Daniel, Allo, Patrick, Taddeo, Mariarosaria, Wachter, Sandra, Floridi, Luciano, "The Ethics of Algorithms: Mapping the Debate", "Big Data & Society", vol. 3, n. 2, 2016, pp. 1–21.
- Portinaro, Pier Paolo, "Il principio disperazione. Tre studi su Günther Anders", Bollati Boringhieri, Torino 2003.
- Portinaro, Pier Paolo, "Günther Anders: un pensiero senza dimora", prefazione a G. Anders, "Essere o non essere. Diario di Hiroshima e Nagasaki", Einaudi, Torino 1961, pp. V–XXX.
- Russell, Stuart, "Human Compatible: Artificial Intelligence and the Problem of Control", Viking, New York 2019.
- Searle, John Rogers, "Minds, Brains, and Programs", "Behavioral and Brain Sciences", vol. 3, n. 3, 1980, pp. 417–424.
- Stiegler, Bernard, "La société automatique. 1. L'avenir du travail", Fayard, Paris 2015.
- Turco, Richard P., Toon, Owen B., Ackerman, Thomas P., Pollack, James B., Sagan, Carl, "Nuclear Winter: Global Consequences of Multiple Nuclear Explosions", "Science", vol. 222, n. 4630, 1983, pp. 1283–1292.
- Turing, Alan Mathison, "Intelligent Machinery" (1948), in D. C. Ince (a cura di), "Mechanical Intelligence", North-Holland, Amsterdam 1992.
- Turing, Alan Mathison, "Computing Machinery and Intelligence", "Mind", vol. 59, n. 236, 1950, pp. 433–460.
- Turing, Alan Mathison, "Macchine calcolatrici e intelligenza", tr. it., in "Menti e macchine", a cura di C. Fabris, Bollati Boringhieri, Torino 2000.
- Turing, Alan Mathison, "Intelligent Machinery, A Heretical Theory" (1951), in B. J. Copeland (a cura di), "The Essential Turing", Oxford University Press, Oxford 2004, pp. 465–475.
- Winner, Langdon, "The Whale and the Reactor: A Search for Limits in an Age of High Technology", University of Chicago Press, Chicago 1986.

- Zuboff, Shoshana, "Il capitalismo della sorveglianza", tr. it. P. Bassotti, Luiss University Press, Roma 2019 (ed. or. "The Age of Surveillance Capitalism", PublicAffairs, New York 2019).