

**L'INTELLIGENZA
ARTIFICIALE
STORIA, PARADIGMI,
CASI STUDIO E
PROSPETTIVE**

**CORSO DI LAUREA IN ECONOMIA AZIENDALE
ING. PAOLA LAPADULA - UNIVERSITÀ DEGLI STUDI DELLA
BASILICATA**

A.A. 2024/2025

SOMMARIO

- Definizioni e contesto dell'IA
- Evoluzione storica e paradigmi epistemologici
- Modelli computazionali: simbolico, connessionista, ibrido
- Tipologie di IA: debole, forte, superintelligenza
- Casi studio: AlphaGo, GPT-4, Watson, DALL·E
- Impatti sociali, culturali e lavorativi
- Rischi etici e problematiche di bias
- AI Act: regolazione e classificazione dei rischi
- Prospettive future, creatività computazionale

COS'È L'INTELLIGENZA ARTIFICIALE?

- L'IA rappresenta un insieme di discipline e tecnologie volte a costruire sistemi computazionali capaci di replicare (e in alcuni ambiti superare) funzioni cognitive tipiche dell'intelligenza umana.
- Comprende domini come percezione automatica, comprensione del linguaggio naturale, pianificazione, apprendimento automatico e ragionamento.
- Le architetture IA sono progettate per eseguire processi decisionali in ambienti incerti, adattarsi all'esperienza, generalizzare conoscenze e interagire autonomamente con l'ambiente.

PERCHÉ È CENTRALE NELL'ATTUALE CONTESTO SOCIO-TECNOLOGICO?

- L'IA è un asse portante della trasformazione digitale, con impatti trasversali in sanità, industria 4.0, giustizia predittiva, sorveglianza algoritmica, educazione personalizzata.
- I modelli IA informano la costruzione di nuovi servizi, modelli di governance e policy tecnologiche.
- Determina sfide etiche su autonomia decisionale, privacy computazionale, equità algoritmica e accountability automatizzata.

PRINCIPI OPERATIVI GENERALI

- Approccio **simbolico** (top-down): rappresentazione esplicita della conoscenza tramite strutture logiche formali e regole inferenziali.
- Approccio **connessionista** (bottom-up): apprendimento distribuito tramite reti neurali artificiali, ispirate al funzionamento sinaptico-biologico.
- Approccio **ibrido**: integrazione funzionale di conoscenza dichiarativa e apprendimento adattivo.

LINEE EVOLUTIVE DELL'IA

- Origini concettuali: automi meccanici (Herone), logica booleana, macchina di Turing (1936)
- 1956: nascita della disciplina alla conferenza di Dartmouth (McCarthy, Minsky, Shannon)
- Anni '70-'80: fioritura dei sistemi esperti (es. MYCIN, DENDRAL), seguiti da crisi per overfitting epistemologico
- Anni '90-2000: rinascita con tecniche probabilistiche e apprendimento automatico
- 2012: Deep Learning e successo di AlexNet → boom di modelli LLM e multimodali

MODELLI COMPUTAZIONALI E PARADIGMI EPISTEMOLOGICI

- **IA simbolica (GOFAI):** si fonda sull'ipotesi del sistema di simboli fisici (Newell & Simon), con regole logico-deduttive esplicite.
- **IA connessionista:** sistemi sub-simbolici che apprendono pattern dai dati; capacità emergenti non pre-programmate.
- **IA ibrida:** sfrutta la complementarità dei due approcci, ad es. nei sistemi cognitivi artificiali e nella robotica autonoma complessa.

IA SIMBOLICA - CARATTERISTICHE ED ESEMPI

- Utilizza basi di conoscenza e motori inferenziali per il ragionamento automatizzato.
- Esempi storici: ELIZA (elaborazione del linguaggio), MYCIN (diagnostica infettiva).
- Ambiti d'uso: sistemi legali automatizzati, gestione aziendale basata su regole, certificazione e audit.
- Vantaggi: piena tracciabilità, spiegabilità logica, validazione formale.
- Limiti: incapacità di apprendere, vulnerabilità all'ambiguità e incompletezza informativa.

IA CONNESSIONISTA - MODELLI E APPLICAZIONI

- Architetture neurali multilivello addestrate con tecniche di ottimizzazione (es. gradient descent).
- Modelli: CNN (elaborazione visiva), RNN/LSTM (dati sequenziali), Transformer (linguaggio naturale)
- Applicazioni: traduzione automatica, guida autonoma, modelli generativi (GAN, VAE, LLM)
- Criticità: black-boxing decisionale, bias nei dati, elevato costo computazionale, vulnerabilità ad attacchi avversariali

SISTEMI IBRIDI E NEURO-SIMBOLICI

- Integra rappresentazioni simboliche e capacità di apprendimento per affrontare compiti che richiedono astrazione e adattamento.
- Esempi: ACT-R, CLARION, modelli neuro-simbolici (es. IBM Project Debater, robotica semantica)
- Applicazioni in: robotica cognitiva, sistemi esperti di nuova generazione, tutor intelligenti
- Limiti: architetture complesse, difficoltà di ottimizzazione congiunta, compromessi tra explainability e performance

IA DEBOLE, FORTE E SUPERINTELLIGENZA

- **IA Debole (Narrow AI):** orientata a task specifici, non generalizza oltre il dominio di addestramento.
- **IA Forte (AGI):** ipotetica entità in grado di emulare pienamente le capacità cognitive umane, inclusa la coscienza metacognitiva.
- **Superintelligenza:** scenario speculativo in cui l'IA supera l'intelligenza umana in tutti gli ambiti cognitivi, strategici e creativi.
- **Rischi:** disallineamento degli obiettivi, controllo algoritmico, collasso epistemico.

CASO STUDIO – ALPHAGO (DEEPMIND)

- AlphaGo ha segnato un punto di svolta superando giocatori umani nel gioco del Go, noto per la sua complessità combinatoria.
- Architettura: reti neurali convoluzionali + apprendimento per rinforzo + Monte Carlo Tree Search.
- AlphaGo Zero ha eliminato l'addestramento supervisionato, apprendendo da zero tramite auto-gioco.
- Impatto: ridefinizione dell'intuizione computazionale, paradigma della creatività algoritmica.

CASO STUDIO – GPT-4 (OPENAI)

- GPT-4 è un modello multimodale capace di elaborare testo, codice e immagini.
- Basato su architettura Transformer, con miliardi di parametri.
- Addestramento in tre fasi: pre-training, fine-tuning supervisionato, RLHF.
- Capacità: ragionamento contestuale, generazione creativa, superamento di test accademici.

CASO STUDIO – IBM WATSON

- Watson vinse nel quiz televisivo Jeopardy! contro campioni umani grazie a NLP, retrieval semantico e machine learning.
- Architettura: approccio ibrido con più moduli paralleli.
- Applicazioni in sanità, finanza e customer service.
- Limiti emersi: difficoltà di generalizzazione, problematiche di trust, problemi di integrazione.

CASO STUDIO – DALL·E E IA GENERATIVA VISUALE

- Modelli come DALL·E e Midjourney generano immagini a partire da prompt testuali.
- Utilizzano encoder testuale + modello di diffusione (o GAN).
- Impatti: democratizzazione dell'immagine, nuove estetiche, controversie su autorialità e copyright.
- Sfide: deepfake, stereotipi visivi, bias culturali nei dataset.

TRANSIZIONE DA SISTEMI SPECIALISTICI A GENERALI

- Evoluzione da IA focalizzate su task singoli a modelli multitasking e generalisti.
- Dal symbolic AI al foundation model.
- I modelli odierni integrano competenze visive, linguistiche, logiche.
- Potenziale: assistenza cognitiva integrata, robotica autonoma, architetture adattive.

CAMBIAMENTO DI PARADIGMA – DAI DATI ALLE CAPACITÀ EMERGENTI

- Sistemi moderni acquisiscono capacità non previste (es. comprensione simbolica, ironia).
- Fenomeno di scaling laws: maggiore dimensione → maggiore generalizzazione.
- Reti complesse mostrano comportamenti non direttamente tracciabili.

PARADOSSI DELL'INCORPORAMENTO SOCIALE

- Le performance tecniche non garantiscono efficacia sociale.
- Es. Watson in sanità, AlphaGo nel Go, GPT nell'educazione.
- Tensione tra efficienza algoritmica e accettabilità sociale.
- I sistemi IA trasformano anche le aspettative umane.

LAVORO COGNITIVO E RIDEFINIZIONE DELLE PROFESSIONI

- IA modifica radicalmente professioni intellettuali: dal giornalismo alla programmazione.
- Automazione parziale: assistenti, correttori, supporti decisionali.
- Emergono nuove figure: prompt engineer, trainer etico, validatore algoritmico.

DISUGUAGLIANZE ALGORITMICHE E CONCENTRAZIONE DEL POTERE

- Rischio di nuova disuguaglianza digitale: accesso disomogeneo a dati, infrastrutture, competenze.
- Poche entità (Big Tech) controllano sviluppo, risorse e distribuzione dei modelli IA.
- Temi: sovranità tecnologica, sorveglianza, marginalizzazione epistemica.

VERSO UNA REGOLAZIONE RESPONSABILE – AI ACT

- L'AI Act europeo introduce classificazione dei rischi (inaccettabile, alto, limitato, minimo).
- Prevede requisiti di trasparenza, qualità dei dati, tracciabilità, sorveglianza umana.
- Obiettivo: tutelare diritti fondamentali, favorire innovazione sostenibile.

IMPLICAZIONI ETICHE DELL'IA

- Decisioni automatizzate in ambiti sensibili (giustizia, medicina, finanza) sollevano questioni etiche complesse.
- Principi chiave: beneficenza, non maleficenza, autonomia, giustizia, accountability.
- Necessità di rendere i sistemi trasparenti, comprensibili e controllabili.
- Sfida: bilanciare innovazione tecnologica con salvaguardia dei diritti umani.

BLACK BOX E TRASPARENZA ALGORITMICA

- Molti modelli IA (soprattutto deep learning) non offrono spiegazioni accessibili del proprio funzionamento.
- Problema per l'adozione in ambiti regolati: sanità, giurisprudenza, pubblica amministrazione.
- Explainable AI (XAI): metodi per rendere interpretabili le decisioni algoritmiche.
- Tecniche: saliency maps, LIME, SHAP, modelli surrogati.

BIAS E DISCRIMINAZIONI ALGORITMICHE

- I modelli apprendono dai dati: se i dati sono distorti, anche le decisioni lo saranno.
- Bias impliciti nei dataset storici: etnici, di genere, geografici.
- Conseguenze: esclusione, ingiustizie sistemiche, perpetuazione di stereotipi.
- Necessità di audit dei dati, fairness-aware learning e testing continuo.

IA E SORVEGLIANZA

- IA viene usata per riconoscimento facciale, tracciamento comportamentale, social scoring.
- Rischi per privacy, libertà individuale, autonomia.
- Sistemi come Skynet (Cina), Project Maven (USA) hanno suscitato forti critiche.
- Dilemma tra sicurezza e diritti civili.

AI ACT – CLASSIFICAZIONE DEL RISCHIO

- **Rischio inaccettabile:** vietati (es. manipolazione subliminale, scoring sociale)
- **Rischio alto:** requisiti stringenti (es. reclutamento, sanità, giustizia)
- **Rischio limitato:** obblighi di trasparenza (es. chatbot, deepfake)
- **Rischio minimo:** libera circolazione

AI ACT – OBBLIGHI PER I FORNITORI

- Documentazione tecnica, valutazione del rischio, gestione del ciclo di vita.
- Registro pubblico europeo dei sistemi ad alto rischio.
- Obbligo di sorveglianza umana e tracciabilità.
- Sanzioni fino al 6% del fatturato globale in caso di violazioni.

GOVERNANCE E SOVRANITÀ TECNOLOGICA

- L'Europa punta a costruire una “via europea all'IA” fondata su diritti e innovazione etica.
- Contrasto con i modelli di sviluppo centrati su mercato (USA) e sorveglianza (Cina).
- Iniziative: ELISE, CLAIRE, AI-on-Demand Platform.
- Sfida: evitare dipendenza tecnologica e favorire capacità autonoma.

IL RUOLO DELLA FORMAZIONE E DELLA CULTURA CRITICA

- Educazione all'IA come competenza trasversale (scuola, università, pubblica amministrazione).
- Alfabetizzazione algoritmica per cittadini, lavoratori, decisori.
- Cultura critica: capacità di interpretare, controllare, orientare l'uso dell'IA.
- Inclusione dei saperi umanistici nei percorsi STEM.

PROSPETTIVE FUTURE DELL'IA

- Integrazione uomo-macchina (BCI, IA collaborativa, co-creazione)
- IA quantistica e potenziale esplosivo di calcolo.
- Scenari speculativi: AGI, superintelligenza, IA autocosciente.
- Richiamo alla responsabilità collettiva nello sviluppo.

IA E CREATIVITÀ COMPUTAZIONALE

- L'IA viene utilizzata per generare testi, immagini, musica, codice, prodotti di design.
- Tecnologie: modelli generativi (GAN, VAE, modelli di diffusione), LLM, IA evolutiva.
- Dilemma: è davvero creatività o rielaborazione statistica?
- Impatto su arte, editoria, industria culturale.

IA E DIRITTO D'AUTORE

- Chi detiene i diritti sulle opere generate da IA?
- Problemi legali con training su dataset protetti da copyright.
- Modelli normativi emergenti: licenze, watermark, indicazioni di origine.
- Giurisprudenza ancora frammentaria e in evoluzione.

IA E LINGUAGGIO NATURALE – CRITICITÀ E POTENZIALITÀ

- Chatbot, assistenti virtuali, sistemi di traduzione: diffusione massiva.
- Rischi: confusione tra umano e artificiale, diffusione di disinformazione, bias linguistici.
- Potenzialità: supporto linguistico, educazione personalizzata, accessibilità aumentata.

PROMPT ENGINEERING – UNA NUOVA COMPETENZA

- L'efficacia dei modelli linguistici dipende dalla qualità dell'input.
- Tecniche: chaining, role assignment, zero-shot vs few-shot learning.
- Si configura come nuova professionalità tra linguistica, informatica e design dell'informazione.

ESERCITAZIONE – ANALISI DI UN OUTPUT GENERATO DA IA

- Fornire un testo generato da GPT o simile.
- Obiettivo: identificare coerenze, incoerenze, allucinazioni, bias.
- Attività guidata di validazione e riscrittura.

IA E INTERDISCIPLINARITÀ

- L'IA è crocevia tra informatica, neuroscienze, filosofia, diritto, economia, arte.
- Approcci ibridi producono soluzioni più robuste e socialmente accettabili.
- Favorire la cooperazione tra domini per evitare sviluppi tecnocratici.

CONCLUSIONI FINALI

- L'intelligenza artificiale non è solo una tecnologia, ma una forza trasformativa che ridefinisce le relazioni tra uomo, società e macchina.
- Le sue potenzialità richiedono una governance consapevole e interdisciplinare, capace di bilanciare innovazione e responsabilità.
- È necessario un dialogo continuo tra scienza, etica, diritto e cittadinanza attiva per garantire un futuro inclusivo e sostenibile.
- *Quale ruolo vogliamo avere come sviluppatori, ricercatori e cittadini nell'orientare l'IA verso il bene comune?*

TERMINI DELLA LICENZA

- This work is licensed under the Creative Commons Attribution-ShareAlike License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-sa/1.0/> or send a letter to Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.
- Questo lavoro viene concesso in uso secondo i termini della licenza “Attribution-ShareAlike” di Creative Commons. Per ottenere una copia della licenza, è possibile visitare <http://creativecommons.org/licenses/by-sa/1.0/> oppure inviare una lettera all’indirizzo Creative Commons, 559 Nathan Abbott Way, Stanford, California 94305, USA.