
aracne

GERARDO IOVANE

FROM ARTIFICIAL INTELLIGENT AGENTS TO ARTIFICIAL COGNITIVE AGENTS

MODELS, METHODS AND APPLICATIONS

aracne

©

ISBN
979-12-218-2154-3

PRIMA EDIZIONE
ROMA 7 AGOSTO 2025

INDICE

Parte I: Agenti Artificiali e Applicazioni

1. Introduzione agli agenti artificiali

La rapida evoluzione dell'IA sta portando gli agenti intelligenti in ambienti tecnologici complessi. Un agente può essere descritto come un sistema che percepisce, ragiona e agisce autonomamente nell'ambiente per il proprio interesse o per supportare gli altri. Gli agenti intelligenti sono componenti sempre più importanti nei settori della sicurezza informatica, della sanità, della finanza, della robotica e delle città intelligenti. L'implementazione efficiente degli agenti di intelligenza artificiale richiede un'architettura validata e, più in generale, gli agenti intelligenti che possono esibire comportamenti autonomi hanno bisogno di una struttura etica per essere affidabili e replicabili.

Il libro "Agenti intelligenti artificiali: Theory, Code, Applications" descrive le basi concettuali, le implementazioni in codice eseguibile e gli approcci di validazione per gli agenti di intelligenza artificiale. La nostra domanda centrale di ricerca è: come si possono progettare, implementare e integrare eticamente gli agenti intelligenti artificiali per ottimizzare il processo decisionale e l'efficienza operativa in settori ad alto impatto, garantendo la replicabilità e la robustezza attraverso il codice eseguibile e la validazione dei benchmark?

Il pubblico è costituito da ricercatori, professionisti e studenti interessati a un libro che affronti le problematiche di progettazione delle tecnologie ad agenti di intelligenza artificiale. L'obiettivo di questo lavoro è fornire una risorsa unificata per la comprensione dei sistemi ad agenti. L'attenzione si concentra su esempi di codice e applicazioni, metodi di validazione, dimensioni etiche e normative degli agenti nei sistemi critici e progetti riproducibili. Questo libro cerca di mostrare i vantaggi e le carenze della tecnologia ad agenti prendendo in considerazione vari esempi, codice eseguibile e scenari del mondo reale.

Un quadro teorico è presente in tutto il libro. Le basi includono le definizioni di agente, l'evoluzione e i diversi modelli di agente come gli agenti reattivi, cognitivi e ibridi. Includiamo codice in Python utilizzando le librerie Mesa, SPADE e PyTorch per la simulazione, il benchmarking e l'esecuzione di agenti nei settori della sicurezza

informatica, della sanità, della finanza, della robotica e delle città intelligenti. Oltre a supportare l'insegnamento e la prototipazione, si presta attenzione anche alla validazione e alla replicabilità, che sono spesso citate come sfide nella ricerca sugli agenti. I rapidi progressi nella ricerca sugli agenti continuano a sollevare diverse preoccupazioni in merito a etica, pregiudizi, spiegabilità e resistenza a circostanze impreviste. Affrontiamo le preoccupazioni riportate nella recente letteratura sugli agenti, nei benchmark e in altri lavori, come la necessità di standardizzare la valutazione e la validazione e la progettazione di agenti in contesti altamente dinamici.

Il libro si propone di andare oltre l'introduzione ai temi dello sviluppo di sistemi basati su agenti. Si compone di quattro parti. Le sezioni iniziali presentano le definizioni fondamentali, i modelli di agenti, le pietre miliari nella progettazione di agenti, le architetture di agenti e altri principi. Le sezioni successive trattano argomenti più specialistici come l'apprendimento, le sfide etiche e i sistemi multi-agente. Questi capitoli includono quadri teorici per i diversi domini, codice e simulazione. Le restanti sezioni presentano esempi di sistemi basati su agenti con specifiche degli agenti, codice basato su Python e simulazioni di benchmark nei settori della sicurezza informatica, della sanità, della finanza, della robotica e delle città intelligenti. L'ultimo capitolo conclude il libro riassumendone il contenuto e presentando le questioni aperte.

Per fornire un contesto ai contenuti che seguono, il prossimo paragrafo presenta un breve schema della struttura del libro.

1.1 Definizione di agente

L'autonomia, l'attitudine sociale, la reattività e la proattività sono ampiamente comprese come caratteristiche che definiscono un agente nell'intelligenza artificiale, che può essere un sistema software o hardware. Questi attributi gli consentono di percepire l'ambiente, prendere decisioni ed eseguire azioni che corrispondono a obiettivi predefiniti (Wooldridge & Jennings, 1995). Elemento essenziale degli agenti intelligenti, l'autonomia va oltre la semplice assenza di coinvolgimento umano. È la capacità di un sistema di prendere decisioni indipendenti basate su regole interne, obiettivi o esperienze apprese, senza bisogno di una guida esterna costante. Tale autonomia diventa cruciale in

applicazioni vitali come il controllo del traffico aereo o la gestione della rete elettrica, dove è necessario che l'agente funzioni in modo affidabile con poca supervisione. L'abilità sociale, un'altra componente cruciale, facilita la comunicazione, la collaborazione e la negoziazione tra agenti o esseri umani. Questa caratteristica aumenta la funzionalità dei sistemi multi-agente, dove la risoluzione di problemi distribuiti, come nella gestione della catena di approvvigionamento, trae vantaggio dall'intelligenza collettiva. L'adattamento del comportamento in base ai cambiamenti ambientali immediati, noto come reattività, è completato dalla proattività, che dà agli agenti la capacità di perseguire obiettivi a lungo termine attraverso strategie di pianificazione o di previsione. Sistemi dinamici come la gestione del traffico, in cui gli agenti devono trovare un equilibrio tra reattività in tempo reale e intento strategico, dipendono da entrambe le caratteristiche. Gli agenti sono formalizzati come mappature dalla storia delle percezioni alle azioni, chiamate anche "funzioni agente", e questo mette in evidenza i principi matematici alla base della loro progettazione, rivelando anche le difficoltà nel gestire situazioni complesse o ambigue. Le architetture avanzate degli agenti incorporano sempre più spesso capacità come l'apprendimento, l'integrazione della memoria e il ragionamento contestuale per superare queste carenze (Wooldridge & Jennings, 1995). La comprensione e l'applicazione degli agenti intelligenti si basano su questi principi di base, che spingono la ricerca verso sistemi coesivi e modulari che funzionano bene in diversi contesti.

La distinzione principale tra agenti e sistemi software convenzionali risiede nell'autonomia e nell'adattabilità dei primi. Gli agenti intelligenti possono imparare dalle loro esperienze e migliorare le loro strategie nel tempo, consentendo risposte dinamiche ad ambienti complessi e mutevoli, in contrasto con programmi statici limitati da set di istruzioni predeterminati (Chaffer et al., 2024). Ad esempio, nel settore sanitario gli agenti sono essenziali per i sistemi diagnostici, dove migliorano continuamente l'accuratezza attraverso l'analisi dei risultati dei pazienti. Allo stesso modo, nei mercati finanziari, gli agenti di trading modificano le loro strategie in risposta alle tendenze in evoluzione, al fine di massimizzare i rendimenti degli investimenti. L'apprendimento per rinforzo e metodi simili danno agli agenti la capacità di aggiornare in modo indipendente le loro politiche decisionali interne, aumentando la loro resilienza in circostanze in rapido cambiamento (Lin, 1992). Soprattutto in settori instabili come il trading finanziario o l'assistenza sanitaria d'emergenza, dove gli agenti devono riconoscere in tempo reale i cambiamenti contestuali cruciali, questa capacità di adattamento è fondamentale.

Tuttavia, l'introduzione di queste tecnologie auto-migliorative richiede un monitoraggio e una validazione accurati, perché l'adattamento può dare luogo a comportamenti imprevisti. La capacità degli agenti non solo di migliorare le prestazioni, ma anche di aderire agli standard etici e normativi sottolinea quanto siano cruciali meccanismi di supervisione affidabili e architetture robuste (Chaffer et al., 2024).

La ricerca sottolinea il ruolo critico delle capacità sociali e comunicative degli agenti intelligenti, soprattutto nei sistemi decentralizzati. Consentendo agli agenti di condividere dinamicamente le informazioni, negoziare e collaborare, queste qualità promuovono un processo decisionale più solido ed efficiente (Jennings, 1993). I vincoli globali sono affrontati con maggior successo dal comportamento coordinato degli agenti rispetto alle strategie centralizzate, come dimostrato da applicazioni reali come la gestione distribuita dell'energia nelle reti elettriche o l'ottimizzazione del traffico urbano. L'abilità sociale comprende attività come la risoluzione dei conflitti e la costruzione del consenso, oltre allo scambio di informazioni di base. Tali meccanismi sono fondamentali in ambienti con risorse limitate, dove gli agenti devono allocare e dare priorità alle risorse condivise senza gravare eccessivamente su componenti specifici del sistema. Inoltre, le comunità di agenti distribuiti migliorano la robustezza intrinseca, riducendo i pericoli associati a singoli punti di guasto (Jennings, 1993). La ridistribuzione del carico di lavoro è resa agevole dalla delega dei compiti e dalla ridondanza tra gli agenti, ad esempio in caso di guasti al sistema. I linguaggi di comunicazione degli agenti che sono standardizzati, tra le altre strutture di comunicazione, servono come base per queste interazioni, incoraggiando lo sviluppo di comportamenti intelligenti emergenti che rispondono a obiettivi collettivi mutevoli (Wooldridge & Jennings, 1995). La socialità, inoltre, promuove la responsabilità e la trasparenza, particolarmente importanti quando le azioni degli agenti hanno un impatto su numerosi stakeholder o devono rispettare i requisiti normativi. Questa caratteristica garantisce che gli agenti intelligenti non solo migliorino la resilienza del sistema, ma offrano anche risultati verificabili e comprensibili, essenziali per la governance in applicazioni delicate.

L'integrazione di meccanismi di apprendimento, come l'apprendimento per rinforzo, nei sistemi di agenti migliora notevolmente le loro capacità, dando loro la possibilità di migliorare le azioni attraverso cicli di feedback di tipo trial-and-error (Lin, 1992). In compiti decisionali sequenziali in cui gli agenti devono navigare in ambienti incerti e stocastici, come nella navigazione robotica o in scenari di gioco, l'apprendimento per

rinforzo è particolarmente efficace. Gli agenti adattivi hanno dimostrato sperimentalmente la loro efficacia, con quelli che utilizzano tecniche di apprendimento sofisticate che superano le loro controparti statiche. Ad esempio, di fronte all'incertezza ambientale, gli agenti che modificano dinamicamente il loro comportamento esplorativo ottengono risultati migliori rispetto a quelli con strategie prefissate. Tuttavia, difficoltà come la lentezza della convergenza e la sensibilità alle variazioni stocastiche costituiscono ostacoli considerevoli, evidenziando la necessità di tecniche più sofisticate come l'ottimizzazione della strategia adattiva e la regolazione degli iperparametri (Lin, 1992). Inoltre, combinare l'apprendimento con la pianificazione o la guida esterna, come l'uso di "stati di addestramento" esplorativi, promuove processi di apprendimento sicuri ed efficaci, consentendo agli agenti di evitare errori potenzialmente disastrosi. Inoltre, queste tecniche promuovono l'apprendimento collaborativo tra gli agenti, consentendo loro di coordinare in modo efficiente le strategie in contesti multi-agente (Lin, 1992). Questi risultati supportano il crescente consenso sul fatto che la capacità di apprendimento è cruciale per gli agenti che operano in contesti in cui le regole predefinite sono insufficienti per affrontare pienamente la complessità del sistema. Di conseguenza, il continuo perfezionamento di algoritmi di apprendimento scalabili ed efficaci è ancora un'area chiave di ricerca finalizzata a implementazioni pratiche nel mondo reale.

L'incorporazione del ragionamento etico e della spiegabilità negli agenti intelligenti è un argomento di discussione e di indagine costante. I moderni framework sottolineano quanto sia cruciale incorporare i principi morali nelle architetture degli agenti, in particolare quando questi vengono utilizzati in campi ad alto impatto come la sanità e la finanza. Poiché un processo decisionale opaco introduce la possibilità di un disallineamento tra le azioni degli agenti e i valori della società, processi decisionali trasparenti e spiegabili sono essenziali per promuovere la fiducia in questi sistemi (Wooldridge & Jennings, 1995). Strutture di governance a livelli, come ETHOS, sono state proposte per regolare la supervisione in base agli effetti sociali delle azioni di un agente, assicurando che le applicazioni con una posta in gioco significativa siano soggette a un esame normativo più severo (Chaffer et al., 2024). Al di là della conformità tecnica, la base etica richiede l'identificazione e la mitigazione dei pregiudizi, nonché la giustificazione delle scelte fatte di fronte all'incertezza. La mancanza di attenzione a questi aspetti rischia di compromettere l'accettazione istituzionale e pubblica degli agenti intelligenti, in particolare nei campi in cui la sicurezza, la privacy o l'equa allocazione delle risorse sono

cruciali (Chaffer et al., 2024). Ad esempio, i sistemi "black-box" nel settore finanziario o sanitario spesso devono affrontare lo scetticismo dell'opinione pubblica e le resistenze normative, rendendo necessari agenti in grado di spiegare i loro ragionamenti in termini comprensibili agli esseri umani. Per affrontare questi problemi e garantire che gli agenti intelligenti siano non solo efficaci dal punto di vista pratico, ma anche affidabili e degni di fiducia, è necessaria una strategia multidisciplinare che incorpori i progressi in materia di apprendimento, autonomia e responsabilità etica.

In conclusione, l'adattabilità, l'autonomia, la socialità e la reattività sono tutti elementi essenziali per la comprensione di base degli agenti intelligenti e ciascuno di essi contribuisce alla loro funzionalità e applicabilità in ambienti dinamici. La progettazione e l'implementazione di questi sistemi sono ulteriormente complicate dall'incorporazione di principi etici e capacità di apprendimento. Nel loro insieme, queste qualità offrono una base per studiare tipi di agenti più specializzati e il loro utilizzo in campi complessi e di grande impatto.

1.2 Agenti semplici, reattivi, deliberativi e cognitivi

Le basi per la comprensione delle diverse complessità e capacità degli agenti artificialmente intelligenti sono stabilite attraverso l'esplorazione di agenti semplici, reattivi, deliberativi e cognitivi. Gli agenti semplici sono caratterizzati da meccanismi semplici di stimolo-risposta, in cui gli input ambientali sono mappati direttamente su output predefiniti, senza alcuna forma di rappresentazione o memoria interna. Questi agenti sono in grado di eccellere in ambienti statici, deterministici e completamente prevedibili, il che li rende adatti, data la loro efficienza e semplicità, a compiti di base, come il recupero automatico di documenti o il controllo di sensori e attuatori di una linea di produzione (Wooldridge & Jennings, 1995). Tuttavia, un limite significativo si rivela nella loro incapacità di gestire scenari imprevisti o di adattarsi a condizioni mutevoli. Le operazioni sono strettamente limitate a direttive basate su regole; di conseguenza, vacillano rapidamente di fronte a stimoli nuovi o imprevisti, con conseguente fragilità del sistema. Questa rigidità è stata illustrata dalle prime implementazioni di agenti semplici in ambienti controllati, in quanto le deviazioni dalle regole programmate portavano al

fallimento operativo, dimostrando l'utilità limitata di tali architetture al di là di domini limitati e statici. Tuttavia, il comportamento deterministico e la facilità di verifica offrono vantaggi distinti nelle applicazioni critiche per la sicurezza, dove la trasparenza e la prevedibilità sono fondamentali, anche se questa affidabilità va a scapito di una limitata autonomia e flessibilità operativa (Wooldridge & Jennings, 1995).

Un modello operativo più avanzato è dimostrato dagli agenti reattivi, invece, grazie all'incorporazione della capacità di adattamento ambientale in tempo reale. Gli agenti reattivi si affidano a stati percettivi immediati per regolare dinamicamente le loro azioni, a differenza degli agenti semplici, senza l'uso di rappresentazioni complesse o di una pianificazione esplicita. Il funzionamento efficace in ambienti dinamici e parzialmente osservabili, come l'esplorazione robotica o gli scenari di navigazione multi-agente, è reso possibile da questa adattabilità in tempo reale (Wooldridge & Jennings, 1995). Steels ha descritto simulazioni di agenti reattivi che utilizzano architetture a sussunzione in sistemi "Mars explorer", che hanno ottenuto prestazioni quasi ottimali nella raccolta di risorse e nell'evitamento di ostacoli, anche in condizioni difficili e imprevedibili. Questo successo sottolinea la forza dei progetti modulari e gerarchici che consentono robustezza e comportamenti emergenti, in particolare in sistemi come la robotica a sciame, dove le regole locali decentralizzate portano alla risoluzione collettiva dei problemi. L'aggiunta incrementale di comportamenti di livello superiore in grado di scavalcare le risposte riflessive è supportata dall'architettura stratificata degli agenti reattivi, che ne aumenta l'utilità e l'estensibilità (Wooldridge & Jennings, 1995). Nonostante questi vantaggi, tuttavia, gli agenti reattivi sono limitati dalla loro incapacità di ragionare sulle esperienze passate o di anticipare le conseguenze future, il che limita la loro applicazione in domini in cui sono necessarie la pianificazione strategica e la formulazione di obiettivi a lungo termine (Wooldridge & Jennings, 1995).

Queste limitazioni sono affrontate dagli agenti deliberativi attraverso l'utilizzo di modelli simbolici interni e meccanismi di ragionamento esplicito, che consentono loro di pianificare ed eseguire comportamenti a lungo termine e orientati agli obiettivi. La strutturazione delle credenze, dei desideri e delle intenzioni di un agente per migliorare la modularità e la spiegabilità del processo decisionale è esemplificata dal framework Belief-Desire-Intention (BDI). Una rappresentazione della conoscenza dell'ambiente da parte dell'agente è fornita dalle credenze, i suoi obiettivi sono riflessi dai desideri e i suoi piani d'azione sono indicati dalle intenzioni. I domini complessi che richiedono alti livelli

di responsabilità e adattabilità, come la gestione del traffico aereo o la cura dei pazienti, sono particolarmente adatti a questa architettura (Wooldridge & Jennings, 1995). La capacità di coordinare i piani, di adattarsi a vincoli in evoluzione e di ottenere risultati superiori in sistemi distribuiti, come la gestione della rete elettrica, è stata dimostrata dall'impiego pratico di agenti deliberativi (Jennings, 1993). Questi agenti eccellono in ambienti che richiedono reattività a vincoli globali e risoluzione di problemi collaborativi, che metterebbero in crisi architetture più semplici. Tuttavia, l'approccio deliberativo introduce un significativo overhead computazionale, in particolare in ambienti volatili che richiedono una frequente rivalutazione del piano. I compromessi inerenti alla gestione di questa complessità sono stati evidenziati dalla ricerca di Kinny e Georgeff attraverso la categorizzazione degli agenti come "audaci", "normali" o "cauti", a seconda della frequenza con cui i loro piani vengono riconsiderati. Negli esperimenti, gli agenti audaci hanno superato quelli prudenti, soprattutto in ambienti ad alta pressione in cui un'azione rapida e decisa era fondamentale (Jennings, 1993). Questi risultati sottolineano l'importanza di una regolazione specifica del dominio per bilanciare la deliberazione con l'efficienza operativa.

Il paradigma deliberativo viene sviluppato dagli agenti cognitivi attraverso l'incorporazione di competenze tradizionalmente associate all'intelligenza umana, come l'adattabilità emotiva e relazionale, la memoria avanzata e le sofisticate capacità di apprendimento. Le applicazioni che richiedono la collaborazione uomo-agente o l'interazione sociale sfumata sono particolarmente efficaci per questi agenti. Il valore dell'incorporazione di competenze simili a quelle umane negli agenti è stato dimostrato dalla ricerca sull'intelligenza artificiale conversazionale: le competenze cognitive, relazionali ed emotive aumentano la fiducia, il coinvolgimento e la soddisfazione degli utenti (Chandra et al., 2022). I settori in cui l'esperienza dell'utente e il successo operativo sono strettamente legati, come l'assistenza sanitaria e il servizio clienti, sono particolarmente importanti per tali capacità. La progettazione di agenti cognitivi, tuttavia, richiede un delicato equilibrio tra l'efficienza strumentale e la replica di qualità simili a quelle umane. Le prestazioni o l'accettazione da parte degli utenti possono essere compromesse da un'eccessiva enfasi su un aspetto a scapito dell'altro (Chandra et al., 2022). La fiducia svolge un ruolo di mediazione in questo equilibrio, suggerendo che gli agenti cognitivi devono dare priorità alla spiegabilità e alla trasparenza per costruire la fiducia degli utenti e mantenere l'allineamento funzionale con i valori della società. Gli

agenti cognitivi devono anche incorporare quadri di apprendimento robusti in domini regolamentati e ad alto rischio per adattarsi ad ambienti in evoluzione, garantendo al contempo equità e responsabilità. L'integrazione di ragionamento simbolico, apprendimento automatico e meccanismi human-in-the-loop è fondamentale per affrontare efficacemente queste sfide (Chandra et al., 2022).

L'adattabilità degli agenti reattivi e cognitivi è notevolmente migliorata grazie all'integrazione di strategie di apprendimento per rinforzo, che consentono loro di ottimizzare il comportamento attraverso un feedback di prova ed errore. I compiti decisionali sequenziali in cui prevalgono incertezze e condizioni stocastiche, come la navigazione robotica o l'allocazione delle risorse in ambienti dinamici, sono particolarmente efficaci per l'apprendimento per rinforzo (Lin, 1992). Gli studi empirici hanno dimostrato che gli agenti con apprendimento per rinforzo superano i progetti statici grazie all'affinamento dinamico delle loro strategie in base alle interazioni ambientali. Gli agenti dotati di apprendimento per rinforzo, ad esempio, possono modificare il loro comportamento esplorativo per adattarsi meglio a scenari rari o inaspettati, ottenendo così un successo superiore. I requisiti computazionali di questi algoritmi, così come le sfide come la lenta convergenza e la sensibilità alle variazioni stocastiche, sottolineano tuttavia la necessità di innovazioni, come la regolazione degli iperparametri e l'ottimizzazione delle strategie adattive (Lin, 1992). L'adattabilità degli agenti di apprendimento non avviene a costo di un disallineamento etico o di comportamenti indesiderati, inoltre, grazie all'incorporazione di salvaguardie come ambienti di formazione esplorativi e meccanismi di supervisione etica. Nonostante questi vantaggi, l'apprendimento per rinforzo introduce rischi di opacità, in quanto gli agenti possono sviluppare strategie difficili da interpretare per gli osservatori umani. La ricerca continua di metodi che bilancino l'adattabilità con la spiegabilità è necessaria per questo motivo, soprattutto in applicazioni delicate, come quelle sanitarie o finanziarie (Lin, 1992).

Una traiettoria evolutiva è evidenziata dalla progressione da agenti semplici ad agenti cognitivi, guidata dalla crescente complessità, capacità e applicabilità del sistema. L'efficienza e il comportamento deterministico sono le priorità degli agenti semplici, ma manca l'adattabilità. La reattività in tempo reale è introdotta dagli agenti reattivi, mentre la pianificazione strategica e il perseguimento degli obiettivi sono consentiti dagli agenti deliberativi. L'intelligenza simile a quella umana e l'adattabilità sociale sono incorporate,

infine, dagli agenti cognitivi, facendo progredire il campo. Queste architetture costituiscono la base per la progettazione di agenti artificiali in grado di affrontare sfide diverse in domini dinamici e ad alto impatto. Il continuo sviluppo dell'integrazione dell'apprendimento, della trasparenza e della responsabilità etica sarà essenziale per liberare il pieno potenziale degli agenti artificiali nelle applicazioni reali.

1.3 Architetture classiche: Funzione agente, PEAS, ambiente

Le architetture ad agenti classiche sono le basi per progettare e valutare il comportamento degli agenti intelligenti all'interno del loro ambiente. Questi approcci, con componenti come la Funzione Agente, il framework PEAS e la tipologia di ambiente basato sugli agenti, forniscono rappresentazioni e semplificazioni fondamentali dei sistemi intelligenti. La Funzione Agente è una mappatura che mappa la storia delle percezioni in azioni. Questa astrazione di un sistema come agente rende più semplice l'analisi di un'entità con uno scopo, in quanto fornisce un modo più semplice per tracciare e prevedere il comportamento dell'agente a partire dagli input (Wooldridge & Jennings, 1995). Tuttavia, a causa della sua semplificazione matematica e dell'approccio diretto, il modello non affronta ambienti non markoviani in cui le percezioni precedenti hanno grande importanza e non c'è integrazione di ragioni contestuali, memoria o pianificazione in più fasi. La mancanza di memoria, anche per l'ultima percezione, limita le capacità dell'agente. Alcuni esempi di applicazione sono la gestione distribuita dell'elettricità e la cura dei pazienti (Wooldridge & Jennings, 1995). In questi due ambiti, la Funzione Agente può scomporre le interazioni agente-ambiente con determinati input e output, ma poiché l'agente non considera le priorità globali e ha un obiettivo fisso o un compito da svolgere in ogni situazione, può essere meno che perfetto. In un caso di applicazione realistica di questi domini, gli agenti non potrebbero funzionare bene in assenza di memoria, a causa della natura fluttuante delle operazioni. Ad esempio, un agente per l'elettricità potrebbe operare senza considerare l'effetto sulle altre macchine della rete o una situazione di rete dinamica, per cui anche se ogni macchina sta operando al suo massimo livello di produttività, potrebbe operare contro l'intera rete. L'evoluzione dei sistemi intelligenti richiede l'ibridazione dei sistemi, in cui la rappresentazione matematica e la

semplificazione sono abbinate ad altri moduli (Jennings, 1993; Wooldridge & Jennings, 1995).

Il framework PEAS (Performance measure, Environment, Actuators, and Sensors) è utile per specificare il compito e definirne gli obiettivi. Fornisce una metodologia per definire e implementare sistemi intelligenti. La prestazione è il modo in cui un agente misura il suo successo e il grado di successo dell'obiettivo, ed è determinata da standard di prestazione esterni o da valori intrinseci. L'ambiente rappresenta il luogo in cui l'agente opererà e limita o fornisce risorse all'agente. Gli Attuatori determinano le possibili azioni che un agente può eseguire nel suo ambiente e i Sensori descrivono quali percezioni un agente può ricevere dal suo ambiente (Wooldridge & Jennings, 1995). Poiché questa architettura consente una rappresentazione più formale dei compiti in modo semplice, la specifica può essere migliorata testando o modificando ciascuno dei suoi componenti attraverso diverse interazioni e casi d'uso. Ne è un esempio l'applicazione per il controllo del traffico aereo e l'ottimizzazione degli ospedali, dove il PEAS semplifica la comprensione degli scenari applicativi, consentendo la progettazione di agenti intelligenti con prestazioni migliori (Wooldridge & Jennings, 1995).

I limiti di questo metodo sono che in un ambiente molto complicato, dove le definizioni dei compiti sono vaghe o incerte, la caratterizzazione delle prestazioni, dell'ambiente, degli attuatori e dei sensori potrebbe essere difficile. Questo potrebbe far sì che le misure di prestazione e l'ambiente non siano ottimali nella specifica del compito, facendo sì che l'agente non operi al massimo delle sue potenzialità. Per un approccio migliore, l'architettura deve incorporare sistemi di misurazione delle prestazioni che si adattino a diversi scenari, dove, utilizzando modelli di apprendimento automatico, gli agenti possono essere ottimizzati in base ai cambiamenti dell'ambiente (Wooldridge & Jennings, 1995).

L'ambiente degli agenti deve essere ben compreso per un migliore sviluppo. Le distinzioni tra completamente e parzialmente osservabile, deterministico e stocastico, episodico e sequenziale, statico e dinamico, discreto e continuo hanno un grande impatto sull'architettura degli agenti.

In un ambiente completamente osservabile, gli agenti possono avere una mappatura più diretta dei valori dei loro sensori sugli attuatori, ma per funzionare in ambienti parzialmente osservabili è necessario sviluppare un modello decisionale più complesso. In questo caso, gli agenti devono considerare la probabilità o la storia per agire in modo

da garantire la massima probabilità di successo. Questo concetto di processo decisionale probabilistico è utile anche per gli ambienti stocastici, dove esiste più di un risultato per un'azione, anche con una condizione precedente nota. Ad esempio, per uno scenario che consiste nell'apertura di valvole all'interno di un impianto di produzione, se l'ambiente in cui operano è parzialmente osservabile, allora devono tenere traccia delle loro azioni precedenti per eseguire azioni future, considerando che la probabilità di una perdita è maggiore se alcune valvole sono chiuse. Inoltre, questo compito può essere considerato stocastico perché la chiusura di una valvola potrebbe avere esiti multipli, in quanto potrebbe chiudersi completamente o non chiudersi abbastanza per mantenere la pressione. Un altro modo di descrivere il compito dell'esempio precedente è episodico o sequenziale. Per gli ambienti episodici, l'agente riceve percezioni solo su una singola istanza nel tempo e deve agire solo su quella condizione per l'intero obiettivo. Negli ambienti sequenziali, come nell'esempio delle valvole dell'impianto, ci sono diverse istanze di input e azioni che l'agente deve eseguire per raggiungere il suo obiettivo, quindi richiede una maggiore intelligenza per operare in modo efficiente (Wooldridge & Jennings, 1995). Infine, se le condizioni delle valvole nell'esempio precedente possono essere considerate statiche, in cui i fattori esterni o gli eventi che possono cambiare le loro condizioni sono ignorati o eliminati, o dinamiche, allora possono anche essere considerate discrete (valvola aperta o chiusa) o continue (valvola leggermente aperta), facendo sì che gli agenti abbiano un insieme quasi infinito di possibilità.

L'intelligenza dell'agente deve evolversi in qualcosa di più robusto per svolgere compiti più complicati in modo efficace ed efficiente (Wooldridge & Jennings, 1995).

Le strutture gerarchiche e modulari sono un passo fondamentale per la soluzione di problemi di controllo complessi. L'aspetto gerarchico consente di semplificare la complessità di ciascun livello. In base alla scala temporale, i livelli superiori eseguono una pianificazione più astratta, mentre i livelli inferiori gestiscono l'esecuzione di azioni su intervalli di tempo più brevi. Poiché i livelli non sono collegati e non hanno comunicazioni proprie, non c'è nemmeno la possibilità che un sistema crei un collo di bottiglia perché dipende da un altro che termina i suoi calcoli prima di poter fare i propri. Un altro aspetto importante delle strutture gerarchiche è la modularità, che consente esecuzioni parallele per calcoli più veloci. Questi moduli sono spesso classificati in basati su regole e basati sull'apprendimento, per avere maggiore semplicità e adattabilità. Nei sistemi basati su regole, gli agenti hanno sempre le stesse prestazioni, indipendentemente dal numero di