COMANDO PER LE OPERAZIONI IN RETE
Reparto Operazioni Cibernetiche

REQUISITO TECNICO OPERATIVO

Relativo a

Acquisizione di un sistema di integrazione multiruolo per modelli di Intelligenza Artificiale generativi denominato "*Intuitive DEcision Artificial intelligence Stack* (IDEAS)"

Edizione Dicembre 2024

INDICE

1. PREMESSA

Sulla base degli attuali ritmi di sviluppo *software* si ravvede la necessità dell'acquisizione di strumenti e capacità algoritmiche sempre più efficaci, che possano potenziare lo sviluppo capacitivo del COR al fine, da un lato, di produrre *tool* che possano rappresentare una innovazione assoluta, dall'altro, di adottare e sviluppare modelli algoritmici in grado di generare ottimizzazione e regolazione dei processi produttivi interni, migliorando l'efficienza.
Il settore dell'intelligenza artificiale, ed in particolar modo quello degli algoritmi generativi, è un settore di frontiera, che offre, già da oggi, varie soluzioni sul mercato rispondenti all'esigenza.

2. SITUAZIONE "*AS IS*"

Attualmente, le operazioni di intelligenza artificiale nel dominio cibernetico soffrono di una frammentazione significativa in termini di integrazione e gestione delle risorse di Intelligenza Artificiale (IA). Sebbene i modelli *Large Language Models* (LLM), le reti neurali convoluzionali (CNN) e le *Generative Adversarial Networks* (GANs) rappresentino tecnologie avanzate con potenziale elevato, la loro combinazione efficace in una piattaforma unificata rimane una sfida. A questo si aggiunge la problematica della perdita di informazioni critiche verso le grandi *corporation* che detengono i modelli migliori e meglio addestrati, con l'inevitabile trasferimento di conoscenza dagli esperti di dominio. Inoltre, l'impossibilità di modificare dinamicamente le configurazioni del sistema limita la capacità di risposta alle minacce cibernetiche, che evolvono rapidamente. La mancanza di un controllo centrale capace di orchestrare le interazioni multimodali tra diversi elementi di IA compromette ulteriormente l'efficacia operativa. In assenza di un *framework* interattivo che sfrutti contemporaneamente la creatività dei modelli di IA e la supervisione umana, molte potenziali soluzioni rimangono inespresse. Per quanto sopra, l'obiettivo del Reparto Operazioni Cibernetiche è, pertanto, quello di acquisire una capacità basata su un sistema di IA ***ON PREMISES*** in grado di:

- combinare in maniera efficace, in una piattaforma unificata, diversi modelli di IA, permettendo così di ridurre notevolmente la frammentazione della gestione e integrazione delle risorse di IA;
- mitigare la problematica della perdita di informazioni critiche verso le grandi *corporation* che detengono i modelli di IA migliori e ben addestrati.

3. SITUAZIONE "*TO BE*"

Al fine di dare corpo a quanto necessario per implementare la citata capacità si rende necessario acquisire una piattaforma basata almeno sui seguenti moduli:

- **Modulo di generazione intuitiva**
  Lo scopo di questo modulo è fornire un sistema che integri le più efficaci tecnologie IA conosciute in letteratura per realizzare una matrice di elementi componibili al fine di ottenere una *pipeline* dinamica di elaborazione generativa che massimizzi le capacità creative ed intuitive pur mantenendo elevata stabilità degli *output*. La *pipeline* di produzione dinamica dovrà essere coordinata da un agente controllore automatizzato (*Intuitive decision-making agent, IDMA*) che regoli le interazioni tra gli elementi e verifica l'attendibilità degli *output*,

scegliendo quali tra le numerose soluzioni proposte dai modelli interni siano quelle più creative, stabili e coerenti. In tale modulo dovrà essere implementata la funzione di composizione dinamica dei modelli di IA ossia, occorrerà sfruttare il concetto della composizione di numerosi blocchetti elementari per la realizzazione di funzioni complesse, con l'aggiunta, in questo caso, di una capacità auto-compositiva dinamica che dovrà permettere al sistema di automodificarsi in base alle esigenze.

Nello specifico, gli elementi che dovranno essere considerati per la costituzione dei blocchi base sono: *Large Language Model* (LLM), *Generative Adversarial Network* (GAN), *Convolutional Neural Network* (CNN), *You Only Look Once* (YOLO), *Single Shot MultiBox, Detector* (SSMD), *Mask* R-CNN.

- ***Multipurpose Decision Array* (MDA)**

Lo scopo del modulo è quello di limitare il rischio di generazione "instabile" degli *output:* pertanto la generazione stessa dovrà avvenire sotto il controllo di un elemento decisionale chiamato *Multipurpose Decision Array* (MDA) capace di validare le risposte di LLM multipli ed estrarre quelle più coerenti con la richiesta posta dall'utilizzatore. All'interno del MDA dovrà essere previsto un analizzatore dello spazio delle decisioni in grado di valutare la moltitudine di soluzioni proposte dai diversi modelli, valutando quelle che minimizzano la distanza con il quesito posto. In particolare dovrà implementare soluzioni di:

a) *Decision Tuning* che dovrà far variare dinamicamente la soluzione proposta, rimettendola in loop, al fine di misurarne la stabilità e favorire l'eventuale fuoriuscita da possibili minimi locali;
b) *Decision Scoring* dovrà assegnare le metriche corrette (selezionate da uno spazio delle metriche) ed assegnare dei punteggi *score* sulla base dei modelli di IA in esame;
c) Il *Decision Selection*, basato su sistema ad apprendimento automatico non supervisionato, dovrà eseguire la *model selection,* tra i tanti modelli di IA che hanno prodotto una risposta al quesito, in funzione dei parametri ricavati dinamicamente dal quesito stesso.

- ***Resource Matrix***

Tale modulo dovrà implementare la matrice delle risorse di AI contenente gli insiemi di modelli pre-addestrati che possono essere singolarmente gestiti con tecniche di *fine-tuning* e nel caso di LLM anche con tecniche RAG (*Retrieval-augmented generation*) o anche più recenti tecniche *Fine-tuning Augmented Embedding* (FAE).

- ***Adversarial Model Loop Generator***

Tale modulo dovrà permettere la generazione di modelli in opposizione di fase allo scopo di enfatizzare creazioni di contenuti creativi ed al contempo stabili e convincenti.

- ***Intuitive Decision Making Agent***

Tale modulo dovrà fornire una capacità decisionale per tutti i modelli utilizzati e di coordinamento delle azioni multimodali mediante le seguenti funzioni:

- la composizione della *pipeline* iniziale (*Control Composer*);
- l'analisi della stabilità della composizione;

- analisi della temperatura (intesa nel senso di generazione entropica) per l'enfatizzazione della creatività delle soluzioni;
- analisi della realizzabilità delle soluzioni proposte mediante un modello critico di confronto ed opportune metriche;

- ***Multipurpose Decision Array***.

Il modulo ha lo scopo di decidere il dosaggio del contributo dei singoli modelli di IA, intervenendo quindi sulla composizione numerica e tipologica complessiva dei modelli di IA da utilizzare.

- ***Pipeline Generator***

IDMA e MDA dovranno essere tra loro interoperabili e dovranno costituire il *pipeline generator* che ha lo scopo complessivo di produrre *output* coerenti dalla composizione parallela di modelli e di *loop* di modelli. Gli output saranno proposti, mediante una interfaccia grafica all'utente, il quale, sulla base dell'utilizzo, genera dei *feedback* (espliciti o impliciti) che vengono poi riportati al modello per ulteriori affinamenti.

Il sistema complessivo viene rappresentato dalla seguente figura:

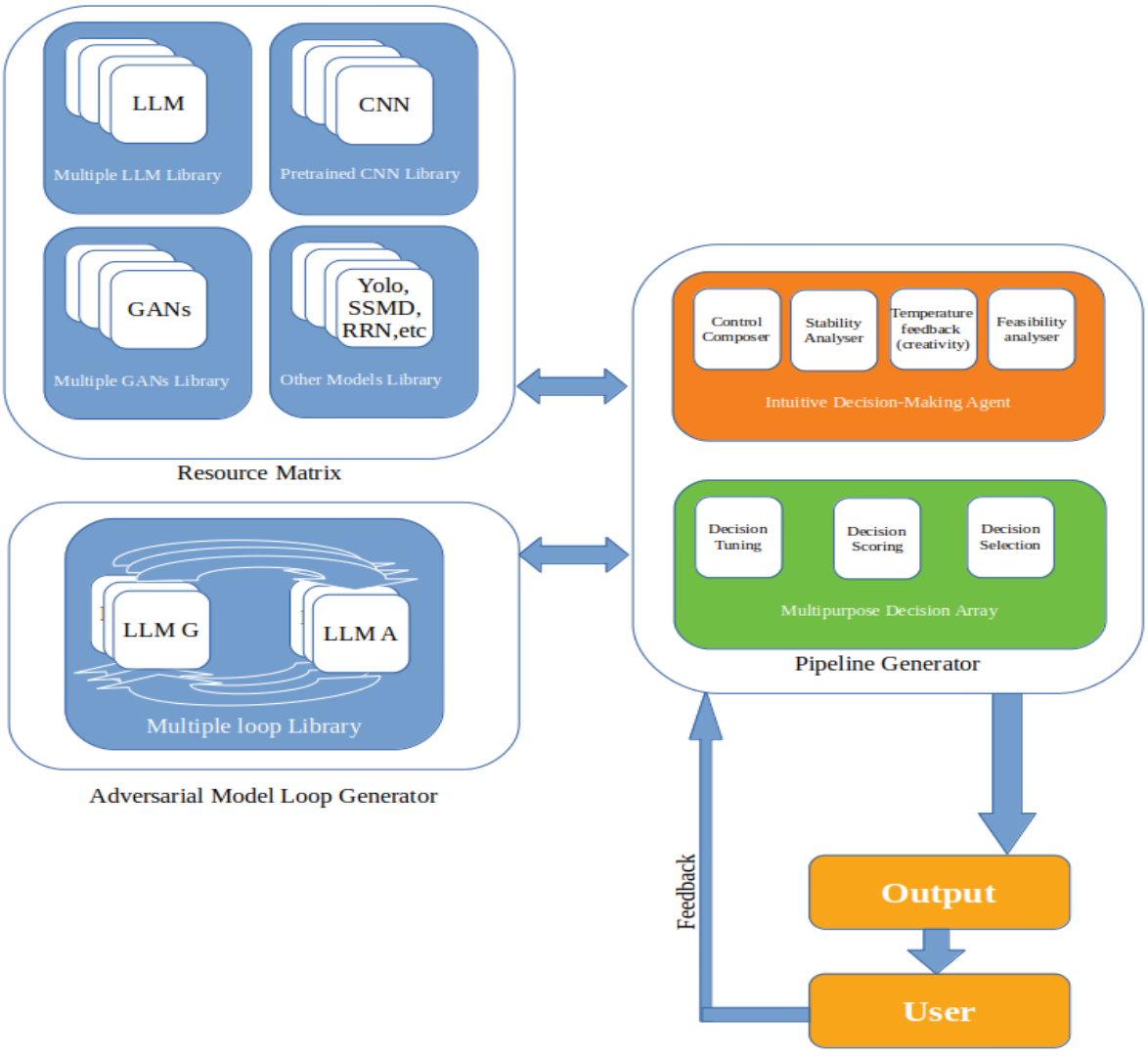
Descrizione dell'integrazione tra i sistemi

Il sistema dovrà permettere di cambiare radicalmente l'approccio ai modelli AI integrati, implementando una piattaforma basata sul controllore intuitivo (*Intuitive Decision-Making Agent* o IDMA). Questo sistema fungerà da cervello centrale, coordinando l'interazione tra diversi modelli AI attraverso *pipeline* dinamiche in modo ottimizzato e coerente. L'IDMA combinato con il *Multipurpose Decision Array* (MDA) creerà una matrice decisionale robusta per validare e ottimizzare gli *output* generati da multiple sorgenti AI.

Il modello proporrà l'uso di una *Resource Matrix* che includerà una vasta gamma di elementi pre-addestrati combinati attraverso tecniche di *fine-tuning* e di generazione aumentata. Questo garantirà risposte efficaci e su misura per compiti specifici, migliorando la stabilità e la risposta innovativa delle soluzioni proposte dal sistema. Dovrà inoltre essere incluso un *Adversarial Model Loop Generator*, che bilancerà la creatività e la stabilità, producendo contenuti innovativi e sostenibili.
Il risultato sarà una *pipeline* generativa capace di confrontarsi con minacce dinamiche in tempo reale, mantenendo il controllo sulle decisioni critiche con il *feedback* degli operatori umani. La creazione di una *pipeline* dinamica in cui numerosi modelli di intelligenza artificiale interagiscono per formare un sistema automorfico rappresenta l'elemento fondante del progetto. In tale contesto lo sviluppo di un sistema automorfico dovrà garantire le seguenti caratteristiche:

- **integrazione semantica:** capacità avanzate di comprendere e utilizzare informazioni in modo semantico tra modelli, migliorando la precisione delle risposte;
- **apprendimento trasferibile:** sistemi che possono trasferire con successo conoscenze tra contesti e applicazioni diverse, riducendo i tempi e i costi di addestramento;
- **Cooperazione Sinergica:** modelli progettati per lavorare insieme in modo efficiente piuttosto che in isolamento;

- **ottimizzazione multimodale:** processi che integrano e utilizzano vari tipi di *input* (testo, immagini, audio) per una risposta dettagliata e precisa;
- **regolazione automatica:** capacità intrinseche dei modelli di adattarsi dinamicamente a *input* variabili mantenendo prestazioni elevate;
- **resilienza ai fallimenti:** protocolli di recupero automatici per mantenere l'operatività continua nonostante eventuali guasti;
- **modularità scalabile:** un'architettura costruita modularmente per la facilità di scalabilità e aggiornamento;
- **meta-apprendimento:** capacità dei modelli di apprendere come migliorare continuamente il proprio processo di apprendimento;
- **prevenzione dei Bias:** algoritmi dedicati alla gestione e alla riduzione dei *bias* nei dati e nelle risposte;



4. GAP ANALYSIS

Sulla base dell'analisi fatta si rende pertanto necessario porre in acquisizione i seguenti elementi:

Acquisizione sistema IDAS ON PREMISES	*Costo*
▪ **1^ FASE:** ○ **T3=T0+3 Mesi** per la preparazione iniziale destinata all'acquisizione delle risorse già orientate e predisposte per poi essere utilizzate su HPC della DIFESA (*High Power Computer*), collezionamento delle ricerche e *papers* scientifici pubblici e studio degli stessi, definizione delle specifiche di progetto; **Fornitura:** Analisi dello stato dell'arte dei pertinenti modelli e definizione delle specifiche di progetto. **Tipo di fornitura:** Rapporto tecnico ○ **T6=T3+3 Mesi** per primo sviluppo matematico dei modelli, raccolta dati, processamento dei *dataset*, sviluppo prime metriche di modello, sviluppo basi teoriche per *l'Intuitive Decision-making Agent (IDMA)*. **Forniture:** Primo modello di test con elementi di verifica delle capacità automorfiche con verifica di esempi appropriati delle capacita decisionali, *dataset* funzionali, metriche operative. **Tipo di fornitura:** Rapporto tecnico ○ **T09=T6+3 Mesi** per la Modellizzazione *alpha* del controllore decisionale IDA, primi addestramenti, primi *test*. **Forniture:** Modello in versione *alpha* addestrato (IDA *alpha*), analisi *performance* e prime verifiche. **Tipo di fornitura:** Dimostratore tecnologico (*software*)	**€ 278.160,00 Euro IVA inclusa)**

Acquisizione sistema IDAS ON PREMISES su HPC	*Costo*
▪ **2^ FASE**: **OPZIONALE (a condizione che la fase 1 abbia ottenuto i risultati desiderati) (PROCEDURA DI GARA DA STABILIRE)** ○ **T3=T0+3Mesi** per la creazione protocollo di generalizzazione modelli eterogenei e matrice delle risorse dei modelli. Creazione e resa operativa la matrice delle risorse ed i relativi protocolli di interconnessione al sistema automorfico. Realizzazione dell'infrastruttura capace di generalizzare i modelli, dalla collezione di modelli, e renderli utilizzabili dall'automorfismo. ***Forniture:*** Versione *alpha* della matrice delle risorse e protocolli di generalizzazione dei modelli. ***Tipo di fornitura:*** Dimostratore tecnologico *(software) su HPC.* TRL4 (partenza)/TRL5 (arrivo) ○ **T6=T3+3Mesi** sviluppo protocollo di astrazione LLM in opposizione di fase Alpha dell'*Adversarial Model Loop Generator*. Realizzazione del sistema di *loop* degli LLM avversari, considerando le tecniche per il mantenimento della stabilità, enfasi delle soluzioni originali, riduzione delle allucinazioni. ***Forniture:*** Versione alpha modulo *Adversarial Model Loop Generator*, protocolli di generalizzazione dei modelli. ***Tipo di fornitura:*** Dimostratore tecnologico (software) su HPC. TRL4/TRL4 ○ **T10=T6+4Mesi**: realizzazione del modello MDA, *Decision Tuning, Decision Scoring, Decision Selection*, prototipo per HPC, Analisi/*test*. ***Forniture:*** Prototipo del MDA comprensivo di analisi e *test*. ***Tipo di fornitura:*** Dimostratore tecnologico (software) su HPC. TRL4/TRL4 ○ **T12=T10+2 Mesi** per integrazione del sistema, analisi coerenza, creazione di agenti di integrazione (tramite API), applicazione alpha per Traduzione codice macchina. ***Forniture:*** Consegna sistema complessivo versione alpha su HPC. Consegna di applicazioni di utilizzo in versione alpha su HPC. ***Tipo di fornitura:*** Dimostratore tecnologico (*software*) su HPC. TRL4/TRL4	

Acquisizione sistema IDAS ON PREMISES su HPC	*Costo*
▪ **3^ FASE: <u>OPZIONALE (a condizione che la fase 2 abbia ottenuto i risultati desiderati) (PROCEDURA DI GARA DA STABILIRE)</u>** o **<u>T3=T0+3Mesi</u>** ottimizzazione dei processi, analisi ottimizzazioni, sviluppo in versione beta del sistema IDA, analisi efficienza protocolli, migliorie, ottimizzazione delle prestazioni ed esecuzione di simulazioni su HPC della Difesa; ***Forniture:*** *database* di interazione per *fine-tuning*, IDA in versione beta, rapporti migliorie. ***Tipo di fornitura:*** Dimostratore tecnologico (*software*) TRL4/TRL5, rapporti operativi. o **<u>T6=T3+3 Mesi</u>** per lo sviluppo dell'applicazione "Formulazione Automatica di Ipotesi di Vulnerabilità e Tecniche di Attacco (FAIVTA)", che fa uso del IDA e del sistema automorfico per analizzare ed attuare potenziali tecniche avanzate di attacco ad un sistema informatico. ***<u>Forniture</u>:*** Applicazione FAIVTA, applicazione per la *cyber intelligence*, rapporti di *test*. ***<u>Tipo di fornitura</u>:*** Dimostratore tecnologico (*software*) TRL4/TRL5, documentazione. o **<u>T10=T6+4 Mesi</u>** l'implementazione, su scenari a cura del ROC nel mondo reale cyber, con una valutazione completa delle tecniche di attacco per fornire indicazioni nel campo operativo. Analisi sistema, ottimizzazione modelli. *Fine-tuning* migliorativi, ottimizzazioni, test operativi. ***Forniture:*** Rilascio sistema in versione 1.0, rapporti migliorie, *test* di utilizzo. ***Tipo di fornitura:*** Dimostratore tecnologico livello TRL5 (partenza)/TRL6 (arrivo)	

NOTE

- La 1^ fase dovrà, inderogabilmente essere posta a collaudo entro il mese di **settembre 2025**. Ciò al fine di poter valutare la prosecuzione del progetto disponendo della tempistica utile a soddisfare i necessari aspetti amministrativi e finanziari.
- Le fasi successiva alla 1^ costituiranno **opzioni** eventualmente e singolarmente esercitabili dalla Difesa subordinatamente alla soddisfacente conclusione della fase precedente (sia dal punto di vista prestazionale che da quello di rispetto della *timeline* vincolante sopra descritta).
- Il progetto dovrà essere accuratamente documento sul piano tecnico e rilasciato accompagnato da appositi manuali d'uso.
- La realizzazione dovrà prevedere interazione continua con il personale del Reparto Operazioni Cibernetiche, sia tramite riunioni in presenza (presso la sede del COR, almeno 1 volta al mese) sia tramite sistemi di comunicazione a distanza.
- La fase 1 del progetto si sovrapporrà di massima al periodo di installazione dell'HPC presso il COR. Essendo dedicata a ricognizione della letteratura scientifica, studi teorici, *concept development*, l'iniziale indisponibilità dell'HPC sarà assolutamente trasparente, ma dovrà comunque tener conto dell'impiego dei prodotti del presente ordinativo sulla citata architettura HPC. Le fasi 2 e 3 del progetto – destinate allo sforzo realizzativo principale – saranno invece destinate all'impiego di tale avanzata unità di calcolo.