

Borsa di studio attivata ai sensi di quanto disposto dal D.M. n. 1061 del 10/08/2021

Titolo del progetto: SAPIENT: Semantic and Automatic Processing of Information about Environment

La borsa sarà attivata sul seguente corso di dottorato accreditato per il XXXVII ciclo:

MODELLI MATEMATICI PER L'INGEGNERIA, ELETTROMAGNETISMO E NANOSCIENZE

Responsabile scientifico: Domenico Vitulano

Area per la quale si presenta la richiesta: GREEN

Numero di mensilità da svolgere in azienda: 6

Azienda: Expert.ai

Il Dipartimento è disponibile a cofinanziare per un importo pari a euro: 7000

Dipartimento finanziatore: DIPARTIMENTO DI SCIENZE DI BASE ED APPLICATE PER L'INGEGNERIA con delibera del 21 settembre 2021

Progetto di ricerca:

La riduzione degli impatti sul cambiamento climatico risulta essere uno dei pillar non solo nell'ambito del PON Ricerca e Innovazione 2014-2020 ma anche in diverse direttive dei decisori politici con l'obiettivo comune di un ambiente ecosostenibile. La valutazione dell'impatto ambientale (VIA) è definita come una procedura di tipo tecnico-amministrativo composta di una serie di fasi obbligate per valutare in via preventiva o correttiva fattori indesiderati in una determinata area. Quando quest'ultima è sufficientemente ampia e articolata, la raccolta di informazioni di tipo ambientale, economico, politico, strategico e sociale diventa un'operazione complessa che richiede la raccolta di tante informazioni, generalmente disponibili nel web. Il processo di automatizzazione di raccolta ed elaborazione di tali informazioni sarà oggetto di questa proposta di Dottorato Innovativo (DI). In particolare, esso si focalizzerà sull'individuazione automatica delle informazioni reperibili sul web. Lasciare, infatti, ad operatori umani tale compito è sicuramente dispendioso (per progetti che coinvolgono regioni ampie diventa addirittura impossibile). Inoltre, tale operazione è molto complicata (talvolta addirittura impossibile) quando la documentazione da analizzare è costituita da pagine nelle quali il testo è riportato su due o più colonne o in tabelle e, ancora peggio, nei casi in cui il testo è scritto di pugno. In tutti questi casi i comuni OCR (Optical Character Recognition), anche commerciali, non consentono una estrazione soddisfacente del testo su cui applicare l'analisi semantica, inficiando l'estrazione automatica della conoscenza d'interesse per il cliente. Il Progetto di DI prevede lo studio, la ricerca e l'implementazione di tecniche di Intelligenza Artificiale seguendo un approccio ibrido di analisi delle immagini e NLU (Natural Language Understanding) per la comprensione semantica di immagini complesse: contenenti testo su più colonne e in tabelle o scritto di pugno. Durante il progetto saranno usate le più innovative tecniche basate sull'uso delle reti neurali profonde: un ottimo candidato potrebbero essere le Convolutional Neural Networks alle quali si pensa, a seconda dei casi, di apportare un contributo in termini di tecniche avanzate di apprendimento percettivo (che risulta essere uno dei maggiori challenge in AI e Machine Learning). Le tecniche basate sulla percezione visiva sfruttano 'i pregi e i difetti' del sistema visivo umano per poter ridurre la complessità del problema permettendo, quindi, di ottenere risultati migliori con un minore costo computazionale.

Il buon esito di questo dottorato di ricerca industriale consentirà di progettare, realizzare e validare in forma prototipale il modulo (SAPIENT), che permetterà una facile e rapida elaborazione di grandi quantità di documenti relativi a informazioni nel settore ambientale.

Gli attori coinvolti saranno Expert.ai, impresa multinazionale da anni leader mondiale di soluzioni di Natural Language Processing (NLP) e il Dipartimento SBAI (Sapienza) che ha un'esperienza consolidata su image processing, tecniche

percettive per l'elaborazione di segnali e immagini con l'uso di reti neurali e machine learning.

In particolare, l'idea di questa proposta nasce dall'esigenza di Expert.ai di estendere le proprie competenze e il proprio mercato al green, dopo numerosi successi in ambito civile e militare nei settori della sicurezza. Il raggiungimento di tale obiettivo potrà avvalersi delle competenze del responsabile scientifico in ambito green ed in particolare: nel remote sensing (monitoraggio della terra, frane, colture etc.), nell'agricoltura di precisione e nell'inquinamento atmosferico indoor (brevetto internazionale PCT/IB2014/064943).

Il Progetto, infine, è in linea con gli obiettivi del Dottorato di Ricerca in Modelli Matematici per l'Ingegneria, Elettromagnetismo e Nanoscienze. Nello specifico, il Progetto condivide le finalità del curriculum Matematica per l'Ingegneria, che prevede di “fornire una solida formazione matematica a giovani ricercatori con una preparazione di tipo scientifico-tecnologico e contemporaneamente dare ai giovani con una preparazione più strettamente matematica un'apertura verso le problematiche provenienti dal mondo produttivo.” Alla luce di queste premesse, il Dottorato in questione incoraggia letesi riguardanti problemi e richieste derivanti da enti esterni. Questi ultimi risultano fondamentali e propedeutici alla formazione completa di un matematico applicato insieme alla capacità di dialogare con il mondo produttivo per lo sviluppo del territorio laziale e nazionale. Non va trascurato, altresì, l'incoraggiamento a svolgere periodi di ricerca presso aziende, enti esterni o istituzioni di ricerca italiane o estere, che necessitino di una modellizzazione matematica avanzata o, più semplicemente, di un uso mirato di metodi e strumenti matematici.

Il cronoprogramma del Progetto sarà:

0-6 mesi: Studio dello stato dell'arte sulle tecniche di valutazione dell'impatto ambientale in diversi contesti

7-12 mesi: Analisi di possibili soluzioni metodologiche (AI e tecniche di segmentazione di documenti)

13-18 mesi: Selezione della soluzione migliore e relativa implementazione prototipale (implementazione in Matlab della soluzione in un contesto scelto)

19-24 mesi: Valutazione sperimentale in laboratorio della soluzione proposta

25-27 mesi: Miglioramento della metodologia proposta sulla base dei test in laboratorio

28-36 mesi: Valutazione sperimentale su dati reali, analisi dei risultati e definizione di eventuali e ulteriori miglioramenti

L'azienda, infine, si impegna a coprire interamente il cofinanziamento previsto (7000 Euro), a garanzia del quale il Responsabile Scientifico mette a disposizione una somma equivalente sui propri fondi (convenzione SBAI-IAC).

Titolo del progetto (inglese): SAPIENT: Semantic and Automatic Processing of Information about Environment

Progetto di ricerca (inglese):