



ALLA MAGNIFICA RETTRICE
DELL'UNIVERSITA' DEGLI STUDI DI MILANO

COD. ID: A022

Il sottoscritto chiede di essere ammesso a partecipare alla selezione pubblica, per titoli ed esami, per il conferimento di un assegno di ricerca presso il Dipartimento di _____ STUDI STORICI _____

Responsabile scientifico: _____ PROFESSORESSA SINEM
ASLAN _____

[Nome e cognome]

CURRICULUM VITAE

INFORMAZIONI PERSONALI

Cognome	FERRO
Nome	SARA

OCCUPAZIONE ATTUALE

Incarico	Struttura
RICERCATORE POST DOC	ISTITUTO ITALIANO DI TECNOLOGIA

ISTRUZIONE E FORMAZIONE

Titolo	Corso di studi	Università	anno conseguimento titolo
Dottorato Di Ricerca	INFORMATICA	CA' FOSCARI	2024
Laurea Magistrale o equivalente	INGEGNERIA DELL'AUTOMAZIONE (LM-25)	UNIVERSITA' DEGLI STUDI DI PADOVA	2017
Master	MACHINE LEARNING E BIG DATA NELLA MEDICINA DI PRECISIONE E NELLA RICERCA BIOMEDICA	UNIVERSITA' DEGLI STUDI DI PADOVA	2018
Esame di stato	ABILITAZIONE ALL'ESERCIZIO DELLA PROFESSIONE DI INGEGNERE DELL'INFORMAZIONE	UNIVERSITA' DEGLI STUDI DI PADOVA	2017



ISCRIZIONE AD ORDINI PROFESSIONALI

Data iscrizione	Ordine	Città

LINGUE STRANIERE CONOSCIUTE

lingue	livello di conoscenza
INGLESE	B2
PORTOGHESE	A1

PREMI, RICONOSCIMENTI E BORSE DI STUDIO

anno	Descrizione premio

ATTIVITÀ DI FORMAZIONE O DI RICERCA

Ho partecipato come TUTOR per il corso di DEEP LEARNING FOR NATURAL LANGUAGE PROCESSING tenuto dal Professore Sebastiano Vascon (A.A. 2023-2024).

ATTIVITÀ PROGETTUALE

Anno	Progetto
2017-2020	Visio System Project (presso la azienda WTM srl - Wire Technology & Machinery a Padova)
2020-2024	Machine learning-supported Historical Document Analysis and Ancient Handwritten recognition - Ph.D. projecft
2023-now	LIBER (Applying Machine Learning and Computer Vision to the Study of Scribal Marks on Cuneiform Tablets) - Collaborator in the Machine Learning part (layout segmentation).
2024-now	RITHMS (Research, Intelligence and Technology for Heritage and Market Security - Collaborator in the Machine Learning part (NLP for extracting entities and relations).



TITOLARITÀ DI BREVETTI

Brevetto

CONGRESSI, CONVEGNI E SEMINARI

Data	Titolo	Sede
05/04/2023	Creating an accurate, well-defined and documented Assyriological Dataset for ML Applications: is it for all?	CAA 2023 (Computer Applications in Archaeology & Qualitative Methods in Archaeology), Amsterdam.
25/06/2023	AI-ASSISTED DIGITALISATION OF HISTORICAL DOCUMENTS	CIPA 2023 Firenze. Documenting, Understanding, Preserving Cultural Heritage: Humanities and Digital Technologies for Shaping the Future.
25/06/2023	CULTURAL HERITAGE DIGITAL PRESERVATION THROUGH AI-DRIVEN ROBOTICS	CIPA 2023 Firenze. Documenting, Understanding, Preserving Cultural Heritage: Humanities and Digital Technologies for Shaping the Future.
15/11/2023	Best Practices in Expanding the Abbreviations. The digitalization of historical abbreviations overlooks "expanded" abbreviations	ClaReNet (Classifications and Representations for Networks) workshop, presso il Römisch-Germanische Kommission (German Archaeological Institute), Francoforte.
15/09/2023	The <i>Specchieri MarVen</i> Dataset: an Abbreviation-rich Dataset in Venetian Idiom	ICIAP 2023 (22 nd International Conference on Image Analysis and Processing), Udine.
06/06/2024	Optical character recognition for ancient manuscripts	RUTE (Revealing Unseen Text with THz Waves) workshop, Venice.

PUBBLICAZIONI

Monografie
[titolo, città, editore, anno...]
[titolo, città, editore, anno...]
[titolo, città, editore, anno...]

Articoli su riviste
Phenomapping of Patients with Primary Breast Cancer Using Machine Learning-Based Unsupervised Cluster Analysis, <i>Journal of Personalized Medicine</i> , MDPI, 2021. Ferro, S., Bottigliengo, D., Gregori, D.,



Fabricio, A. S. C., Gion, M., & Baldi, I.

Boosting CNN-based Handwriting Recognition Systems with Learnable Relaxation Labeling, arXiv, 2024. Ferro, S., Torcinovich, A., Traviglia, A. & Pelillo, M. (under review in Pattern Recognition journal of Elsevier).

Multitemporal Multispectral Dataset for Palaeochannels Segmentation (MAPS), arXiv, 2024. Poggi, G., Yaseen, A., Jaturapitpornchai, R., Ferro, S., Sech, G., Naylor, P., Salvi, M.C., Vascon, S., Le Saux, B., Fiorucci, M., & Traviglia, A.

Atti di convegni

AI-ASSISTED DIGITALISATION OF HISTORICAL DOCUMENTS, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Science, Firenze, 2023.

The Specchieri MarVen Dataset: an Abbreviation-Rich Dataset in Venetian Idiom, ICIAP 2023 (22nd International Conference on Image Analysis and Processing), Udine, 2023.

Exploiting Context in Handwriting Recognition Using Trainable Relaxation Labeling, Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods, Lisbona, 2023.

[titolo, struttura, città, anno]

[titolo, struttura, città, anno]

ALTRE INFORMAZIONI

Esperienza professionale: R&D Computer Vision Engineer (2017-2020)

Nel periodo dal 2017 al 2020 ho lavorato come **R&D Computer Vision Engineer** presso l'azienda WTM s.r.l., occupandomi dello sviluppo di **algoritmi di Computer Vision** per la rilevazione di difettosità su cavi nastrati o cordati prodotti dalle macchine dell'azienda.

Nello specifico, ho sviluppato **algoritmi di visione artificiale** in **C++**, utilizzando principalmente la libreria **OpenCV**, e ho usato **Qt Creator** come Interface Development Environment (IDE), sfruttando le **librerie Qt** per implementare la componente grafica del sistema di visione.

Tra le altre attività:

- Implementazione di librerie per l'**integrazione di SDK proprietari** delle camere, finalizzate all'acquisizione e alla pre-elaborazione delle immagini nell'applicativo.
- Uso di strumenti di **controllo di versione** come **GIT**, con applicativi per la gestione avanzata del codice, tra cui **GitHub**.
- Utilizzo di **Python** e **GPU Nvidia** per ottimizzare il progetto e ampliarne le capacità (ultima attività solo iniziata).

Ho inoltre ricoperto il ruolo di **team leader** per il progetto di visione, occupandomi di assegnare task ai membri del gruppo, coordinare il lavoro e presentare il sistema di visione a potenziali clienti interessati.

Master di II Livello in Machine Learning (2018)

Nel 2018 ho frequentato un Master di II Livello, durante il quale ho approfondito metodi di **Machine Learning supervisionati, semi-supervisionati e non supervisionati**.

Il percorso formativo mi ha permesso di sviluppare:

- Solide basi teoriche sui principali **modelli di Machine Learning**.
- Competenze pratiche nell'implementazione di algoritmi utilizzando il **linguaggio R**.



Questa esperienza ha consolidato le mie capacità di applicare modelli di Machine Learning a problemi reali, combinando teoria e pratica.

Borsa di ricerca: Università degli Studi di Padova (2019)

Nel 2019 ho svolto un'attività di ricerca presso il Dipartimento di Statistica dell'Università degli Studi di Padova, focalizzata sull'analisi delle performance degli studenti nei corsi ad accesso vincolato della Dipartimento di Ingegneria.

Durante il progetto, mi sono occupata di:

- Analisi statistica dei dati utilizzando il **linguaggio R** e l'**IDE RStudio**.
- Creazione di report interattivi e dashboard per la visualizzazione dei risultati mediante **Qlik**.

Questa esperienza mi ha permesso di consolidare le competenze nell'analisi dei dati e nell'uso di strumenti avanzati per l'elaborazione e l'interpretazione statistica.

Borsa di dottorato: Università Ca' Foscari (Venezia) in collaborazione con Centre for Cultural Heritage Technology (CCHT), centro dell'Istituto Italiano di Tecnologia (2020-inizio 2024)

Durante il progetto di dottorato "Machine learning-supported Historical Document Analysis and Ancient Handwritten recognition" ho sviluppato un **modulo di apprendimento** innovativo che sfrutta efficacemente le **informazioni contestuali** durante la etichettatura (trascrizione automatica), un concetto essenziale in Computer Vision e Machine Learning. In particolare, questo nuovo modulo di apprendimento si distingue per la sua capacità di migliorare l'accuratezza della trascrizione automatica di documenti storici senza aumentare la dimensione del modello durante l'inferenza. Inoltre, questo lavoro definisce le **"migliori pratiche"** di **design** di modelli di **Deep Learning** per l'**espansione delle abbreviazioni**, affrontando sfide precedentemente poco esplorate. Durante il progetto di dottorato abbiamo creato e rilasciato un **nuovo dataset** di immagini righe di **documenti storici veneziani** (con relativa trascrizione), primo nel suo genere, per fare l'**addestramento** di modelli di trascrizione automatica. Infine, nella fase finale del dottorato, ho avviato un'interessante analisi riguardante **explainability dei modelli di trascrizione automatica**, che ha fornito spunti preziosi per definire un modello con il **minimo numero di parametri, ottimizzato per il dataset di interesse** (non sono gli iperparametri ad essere ottimizzati, ma è proprio il design architetturale del modello stesso).

Competenze acquisite:

Ho raggiunto una conoscenza molto avanzata del linguaggio di programmazione **Python** e approfondito l'utilizzo di librerie per il DL come **PyTorch** e **HuggingFace**, e consolidato le mie competenze nel **controllo di versione con GIT**, questa volta utilizzando l'istanza di **GitLab** dell'Istituto Italiano di Tecnologia. In questo contesto, ho collaborato con dottorandi e research fellows, avviando pratiche di revisione del codice per migliorare la qualità e imparare reciprocamente. Inoltre, abbiamo introdotto l'uso di test automatizzati per garantire l'affidabilità del software sviluppato.

Ho acquisito esperienza nello sviluppo di **applicazioni multinodo**, sia sfruttando i server locali del centro di ricerca, sia utilizzando il **supercomputer HPC** di Genova.

Sul piano tecnico e teorico, ho maturato una conoscenza approfondita dei modelli di Computer Vision, tra cui **Convolutional Neural Networks (CNN)** e **Transformers**, nonché dei modelli di **Natural Language Processing (NLP)**, come le **Recurrent Neural Networks (RNN)** e le loro varianti, oltre ai **Transformers**.

Post Doc: Center for Cultural Heritage Technology (CCHT), centro dell'Istituto Italiano di Tecnologia (2024-adesso)

Nel progetto a cui sto collaborando, ho acquisito competenze nell'utilizzo e nel fine-tuning di **Large Language Models (LLMs)**, tra cui **LLaMA 3** di Meta e altri modelli di riferimento. Inoltre, ho iniziato a implementare il modello **Retrieval-Augmented Generation (RAG)** per applicazioni avanzate di NLP.



Ho approfondito l'uso di librerie classiche per il Natural Language Processing, come **SpaCy** e **HuggingFace**, e ho sviluppato modelli per il **Named Entity Recognition (NER)** e metodologie automatizzate per la **Relation Extraction (RE)** necessari per la costruzione di **Knowledge Graphs (KG)**.

Attualmente, un paper relativo a questo lavoro è in fase di revisione interna.

Le dichiarazioni rese nel presente curriculum sono da ritenersi rilasciate ai sensi degli artt. 46 e 47 del DPR n. 445/2000.

Il presente curriculum, non contiene dati sensibili e dati giudiziari di cui all'art. 4, comma 1, lettere d) ed e) del D.Lgs. 30.6.2003 n. 196.

RICORDIAMO che i curricula SARANNO RESI PUBBLICI sul sito di Ateneo e pertanto si prega di non inserire dati sensibili e personali. Il presente modello è già precostruito per soddisfare la necessità di pubblicazione senza dati sensibili.

Si prega pertanto di **NON FIRMARE** il presente modello.

Luogo e data: _____ PADOVA _____, 16/12/2024 _____