

**PUBBLICAZIONE, AI SENSI DELL'ART. 19 DEL D.LGS N. 33 DEL 14 MARZO 2013,
MODIFICATO DALL'ART. 18 DEL D.LGS N. 97 DEL 25 MAGGIO 2016 COME
INTEGRATO DALL'ART.1 C. 145 DELLA LEGGE 27 DICEMBRE 2019 N. 160,**

DELLE TRACCE D'ESAME

STABILITI DALLA COMMISSIONE ESAMINATRICE

DELLA SELEZIONE DI SEGUITO INDICATA

NELLA RIUNIONE IN DATA 11/10/2024

TRACCE DELLE PROVE D'ESAME – PROVA ORALE

BANDO N. 400.21 ISTI PNRR

Selezione per titoli e colloquio ai sensi dell'art. 8 del “*Disciplinare concernente le assunzioni di personale con contratto di lavoro a tempo determinato*”, per l'assunzione, ai sensi dell'art. 141 del Contratto Collettivo Nazionale del Lavoro (CCNL) relativo al personale del Comparto Istruzione e Ricerca, Triennio 2019-2021, sottoscritto in data 18 gennaio 2024, di **una unità** di personale con profilo professionale di **Ricercatore III livello**, presso l'Istituto di Scienza e Tecnologie dell'Informazione “A. Faedo” del CNR – sede di Pisa, nell'ambito del PNRR, , Missione 4, “Istruzione e Ricerca” - Componente 2, “Dalla ricerca all'impresa” - Linea di investimento 3.1, “Fondo per la realizzazione di un sistema integrato di infrastrutture di ricerca e innovazione”, finanziato dall'Unione europea – NextGenerationEU – Progetto IR0000013 “SoBigData.it: Strengthening the italian RI for social mining and big data analytics” (CUP B53C22001760006).

BUSTA 1

Domanda 1

Il candidato illustri la propria esperienza di ricerca nell'ambito del “Cross-lingual/cross-modal learning” ed enunci le principali tecniche per la generazione di “word embeddings” presenti in letteratura.

Domanda 2

Il candidato parli della propria esperienza nell'uso di modelli *transformer* e indichi quali *Large Language Model* o *Visual-and-Language Model* ha utilizzato nella sua ricerca.

Domanda 3

Il candidato parli della propria prospettiva di ricerca nell'ambito del “Cross-lingual/cross-modal learning” e su quali aspetti si concentrerebbe nel caso risultasse vincitore della selezione.

Domanda 4

Lettura e traduzione dalla lingua Inglese del seguente testo tratto da: Olof Görnerup and Theodore Vasiloudis, *Mining Similarities and Concepts at Scale*. ERCIM-NEWS 107, September 2016

<https://ercim-news.ercim.eu/en107/special/mining-similarities-and-concepts-at-scale?highlight=WyJkaXN0cmliidXRpb25hbCIsmh5cG90aGVzaXMiXQ==>

“What is similarity? This fundamental, almost philosophical, question is seldom asked in machine learning in all but specific contexts. And although similarities are ubiquitous in the field, they are often limited to specific domains or applications; calculated between users to make recommendations, between websites to improve web searches, or between proteins to study diseases, for example.

To enable similarity and concept mining applicable in a broad range of areas, we have proposed a generalisation of the distributional hypothesis commonly used in natural language processing: Firstly, the context of an object (a word, artist or molecule, for instance) is its correlations (co-occurrences, play order, interactions etc) to other objects. Secondly, similar objects have similar contexts. And thirdly, a group of inter-similar objects form an abstraction – a concept.

In this way, we essentially base the notion of similarity on exchangeability. In the case of language, for example, it should be possible to remove a word from a sentence and replace it with a similar one (consequently belonging to the same concept) and the sentence should still make sense. And the same should be true regardless of the dataset and types of objects, whether it is words, items a user interacts with, people in a social network, or something else.”

BUSTA 2

Domanda 1

Il candidato illustri la propria esperienza di ricerca nell’ambito del “Cross-lingual/cross-modal learning” e spieghi che cosa è il “visual grounding” e come si rapporta con l’area della *distributional semantics*.

Domanda 2

Il candidato parli della propria esperienza nell’uso di modelli transformer e indichi quali sono i modelli *cross-lingual* e *cross-modal* allo stato dell’arte.

Domanda 3

Il candidato parli della propria prospettiva di ricerca nell’ambito del “Cross-lingual/cross-modal learning” e racconti le sfide di ricerca che ritiene più promettenti sulle quali si vorrebbe concentrare nel caso risultasse vincitore della selezione.

Domanda 4

Lettura e traduzione dalla lingua Inglese del seguente testo tratto da: Diego Collarana Vargas (Fraunhofer FIT) and Nassos Katsamanis (Athena RC), *Large Language Models - Introduction to the Special Theme*, ERCIM-NEWS Special Issue, January 2024

<https://ercim-news.ercim.eu/en136/special/large-language-models-introduction-to-the-special-theme>

“These articles highlight the current state of LLM applications and the necessity for robust, Europe-centric research and policy frameworks. These frameworks should be designed to harness the full potential of LLMs while mitigating risks such as digital divides, cultural homogenization, and potential misalignments with European values. Within this context, the transformative impact of LLMs in European education is particularly noteworthy; these technologies, as highlighted in the richest category of this edition, hold the potential to revolutionize learning methodologies, making education more interactive, personalized, and accessible – a crucial advancement in an era where

adapting to rapidly changing skill sets and knowledge bases is increasingly vital. We hope that by building on the perspectives and solutions highlighted in this special edition and by nurturing further research, collaborations, and discussions, we can steer Europe towards a future where the integration of LLMs goes beyond mere technological innovation. These models must align seamlessly with Europe's unique cultural, ethical, and regulatory frameworks, promising a responsible, inclusive, and profoundly transformative digital future for all. We look forward to engaging in discussions about LLMs and ChatGPT as we pursue breakthroughs in Natural Language Understanding (NLU) to develop improved models and applications with a profound and positive impact on society.”

IL PRESIDENTE

Prof. Alessandro Lenci

LA SEGRETARIA

Dott.ssa Barbara Furletti