



PhD in INGEGNERIA DELL'INFORMAZIONE / INFORMATION TECHNOLOGY - 38th cycle

Research Area n. 1 - Computer Science and Engineering

**THEMATIC Research Field: UNDERSTANDING THE LIMITS OF LARGE LANGUAGE
MODELS FOR ASSISTING WITH JOURNALISTIC CONTENT GENERATION AND EVIDENCE
CURATION**

Monthly net income of PhDscholarship (max 36 months)

€ 1400.0

In case of a change of the welfare rates during the three-year period, the amount could be modified.

Context of the research activity

**Motivation and objectives of the research
in this field**

The advent of massive Language Models (LMs) has brought with it the ability to generate high quality and engaging text about topics of public interest. When conditioned on factual sources of information such as evidence from official documents, statistical tables, eyewitness accounts and/or photos of a particular event, the content generation could be used to create journalistic content or assist journalists in the curation and generation of such content. With this use-case in mind, there exist many research questions to investigate:

- How can one combine evidence across different modalities (text, tabular and images) to produce journalistic content?
- How can one prevent the LM from hallucinating facts that are non-deducible from the collected evidence?
- How can one manage the level of paraphrasing to prevent copyright infringement from any of the source documents?
- How can one target content generation at certain type of readership (e.g. sports readers) by tailoring the style of the writing (e.g. with the addition of wit or irony).
- What limits exist (if any) on the reasoning capabilities of Transformer based LMs to combine and make



	<p>deductions from available evidence?</p> <ul style="list-style-type: none"> • Can the same model be used to provide explanations regarding its reasoning over the available evidence?
Methods and techniques that will be developed and used to carry out the research	<p>The PhD student will make the use of Deep Learning techniques such as Transformer-based and related models for conditional text generation, investigating the numerous research questions mentioned above. The PhD scholarship will be funded by ML cube S.r.l., a company providing Artificial Intelligence and Optimization solutions. The PhD student will spend part of their time at the company taking advantage of the wealth of machine, deep and reinforcement learning expertise that exists there. Furthermore, they will apply the results of their research to real industrial use-cases, making use of available data and computing resources that the company possesses.</p>
Educational objectives	<p>The student will learn:</p> <ul style="list-style-type: none"> • research project management and publishing skills, • data science and data analytics skills (particularly deep learning and explainable AI skills), • and presentation and teaching skills, • while developing into a professional researcher.
Job opportunities	<p>Strong employment opportunities exist for PhD level Data Scientists with advanced skills in Deep Learning and experience in Text Processing applications. Companies involved in content generation, journalism and more broadly machine learning will be particularly interested in hiring a PhD-graduate with application experience in this area.</p>
Composition of the research group	<p>0 Full Professors 1 Associated Professors 0 Assistant Professors 2 PhD Students</p>
Name of the research directors	<p>Mark James CARMAN</p>



Contacts

Prof. Mark Carman mark.carman@polimi.it 02 23993623 https://www.deib.polimi.it/ita/personale/dettagli/1439980
--

Additional support - Financial aid per PhD student per year (gross amount)	
--	--

Housing - Foreign Students	--
Housing - Out-of-town residents (more than 80Km out of Milano)	--

Scholarship Increase for a period abroad	
--	--

Amount monthly	700.0 €
By number of months	6

Additional information: educational activity, teaching assistantship, computer availability, desk availability, any other information

Company at which the PhD will perform the external activity

- Name: ML cube S.r.l.
 - Contact: Lorenzo Bisi (lorenzo.bisi@mlcube.com)
 - Planned months: 12 months
 - Activity: Collection of datasets, experimental evaluation on real data available at company.
 - Collaborations: ML cube is itself a spin-off from the Politecnico di Milano and thus there exist many collaborations between the company and a number of researchers at the Politecnico.
- EDUCATIONAL ACTIVITIES (purchase of study books and material, including computers, funding for participation in courses, summer schools, workshops and conferences): financial aid per PhD student
5.707,13 Euro

TEACHING ASSISTANTSHIP: (availability of funding in recognition of supporting teaching activities by the PhD student)

There are various forms of financial aid for activities of support to the teaching practice. The PhD student is encouraged to take part in these activities, within the limits allowed by the regulations.

COMPUTER AVAILABILITY: individual use