



## Opportunity for M.Sc. thesis

<b>Title</b>	Large language models for risk assessment of industrial systems
<b>Motivations and objectives of the research</b>	<p>During the lifetime of industrial systems, accidents and failures threaten the system safety, availability and productivity. Learning from past accidents and failures, by identifying the factors that influence their occurrence and severity, constitutes one of the bases of risk assessment of industrial systems.</p> <p>Textual reports of accidents and failures, compiled by system operators to describe the events, their causes, their consequences, their severities, and, in case of maintenance, the performed inspection, repair and replacement activities, are typically recorded and stored in repositories.</p> <p>Traditionally, this valuable source of information is not systematically exploited in the technical framework of risk assessment. This is mainly due to: <i>i)</i> the extensive use of technical language in the reports, which requires expert-based intervention, <i>ii)</i> the inhomogeneity of the repositories (e.g. due to the subjectivity of the operators writing the reports), and <i>iii)</i> the large number of reports to consider.</p> <p>Large Language Models (LLMs) represent the latest and most promising technological achievements in the field of generative artificial intelligence. They are dialogue systems trained on extensive datasets able to understand language and generate text, and their continuous scaling in dimensions allowed for emergent properties such as reasoning and in-context learning.</p> <p>This thesis work aims at developing a methodological framework based on Large Language Models (LLMs) for systematically exploiting the knowledge, information and data contained in textual reports of accidents and failures in industrial systems for supporting their risk assessment.</p> <p>The effectiveness of the proposed framework is going to be showed considering real repositories of accident and failure reports of industrial assets.</p>

<b>Activities</b>	<ol style="list-style-type: none"> <li>i. Literature review on LLMs and generative models, and their applications to accident and failure reports;</li> <li>ii. Definition of the research problem;</li> <li>iii. Development of the methodological framework;</li> <li>iv. Application of the method to a real case study;</li> <li>v. Validation of the developed framework for supporting risk assessment;</li> <li>vi. Writing of scientific papers</li> </ol>
<b>Possible industrial collaborations</b>	ENI Unit Process Safety
<b>Possible academic collaborations</b>	Centro de Estudos e Ensaaios em Risco e Modelagem Ambiental (CEERMA), Universidade Federal de Pernambuco (UFPE)
<b>Required competencies and skills</b>	<ul style="list-style-type: none"> <li>• Interest in developing innovative applications of Natural Language Processing, Machine Learning and Artificial Intelligence algorithms for risk assessment;</li> <li>• Good knowledge of the topics of risk assessment;</li> <li>• Good knowledge of Python programming or willingness to learn;</li> </ul>
<b>Composition of the research group</b>	<ul style="list-style-type: none"> <li>• Number of Full Professors: 2</li> <li>• Number of Postdoctoral researchers: 1</li> </ul>
<b>Name of the research director</b>	Enrico Zio
<b>Email address</b>	piero.baraldi@polimi.it enrico.zio@polimi.it
<b>Total thesis duration</b>	10 months