




**<SIGHT**

BY SAIPEM



# **A machine learning model for fast identification of technical requirements in project documentation**

**I big data incontrano il manifatturiero:  
opportunità e vantaggi**

**Claudio Nespoli  
Francesco Ramoni**

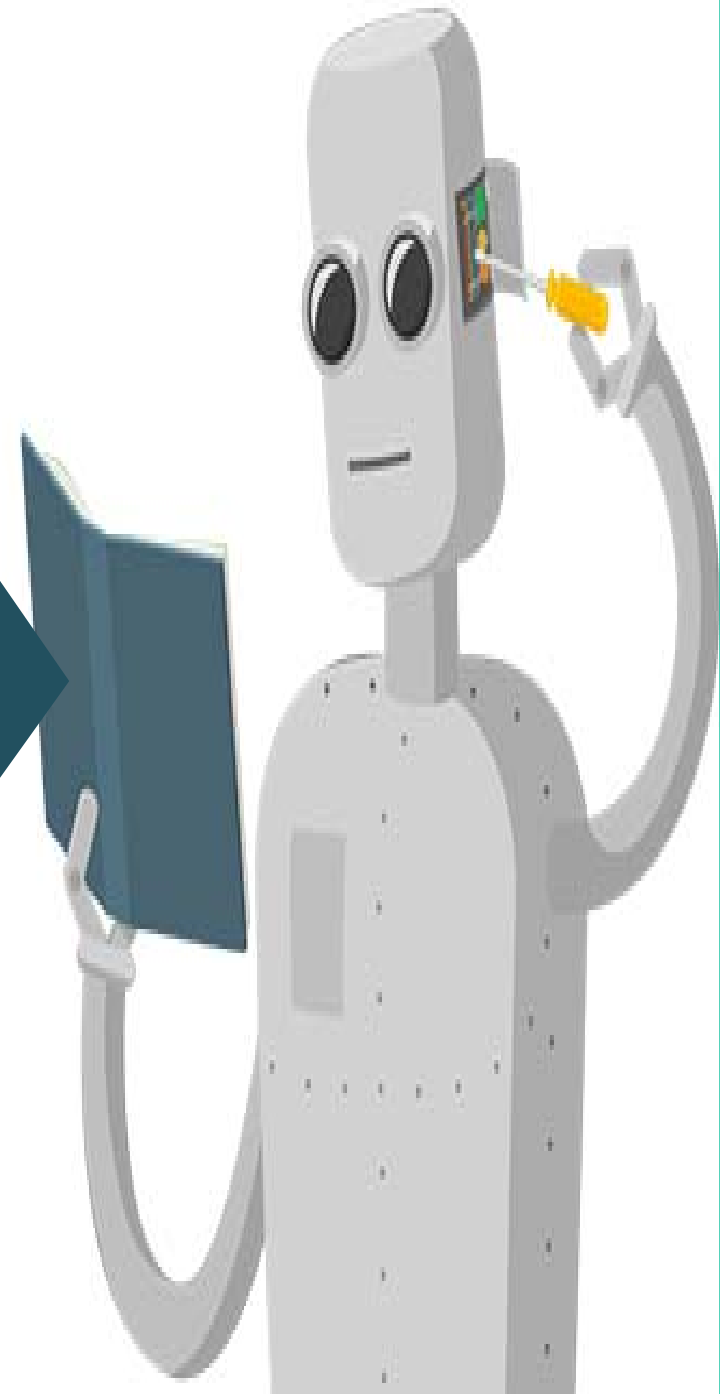
**Cinisello Balsamo – 15/10/2019**

**A machine learning model for fast identification of  
technical requirements in project documentation**

**I big data incontrano il manifatturiero:  
opportunità e vantaggi**

**Claudio Nespoli  
Francesco Ramoni**

**Cinisello Balsamo – 15/10/2019**



# Saipem, a global solution provider

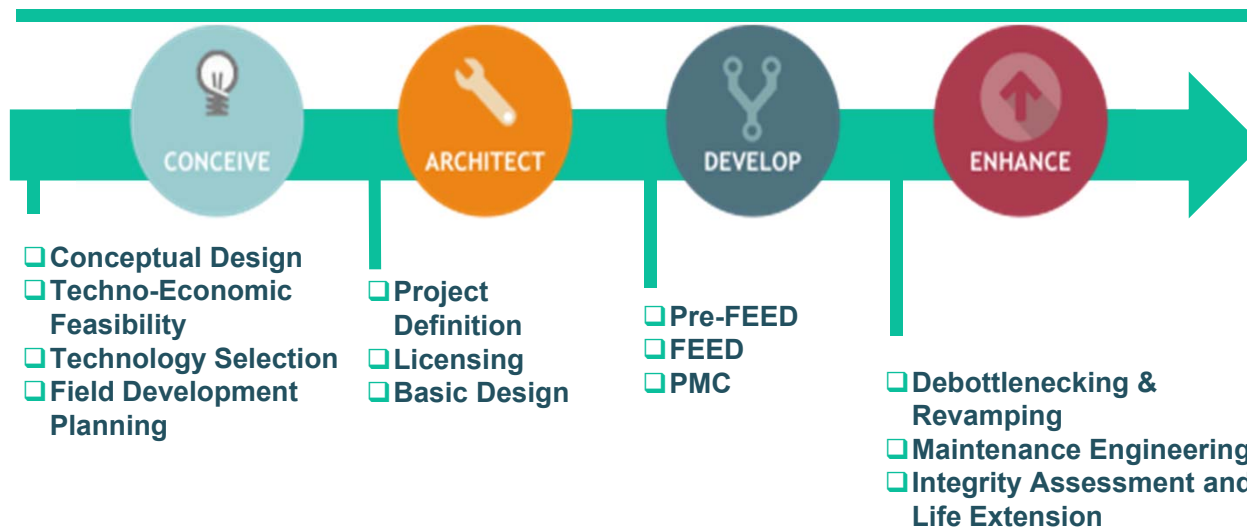
## Business lines overview



# Saipem, a global solution provider

## Xsight division

### Engineering Services



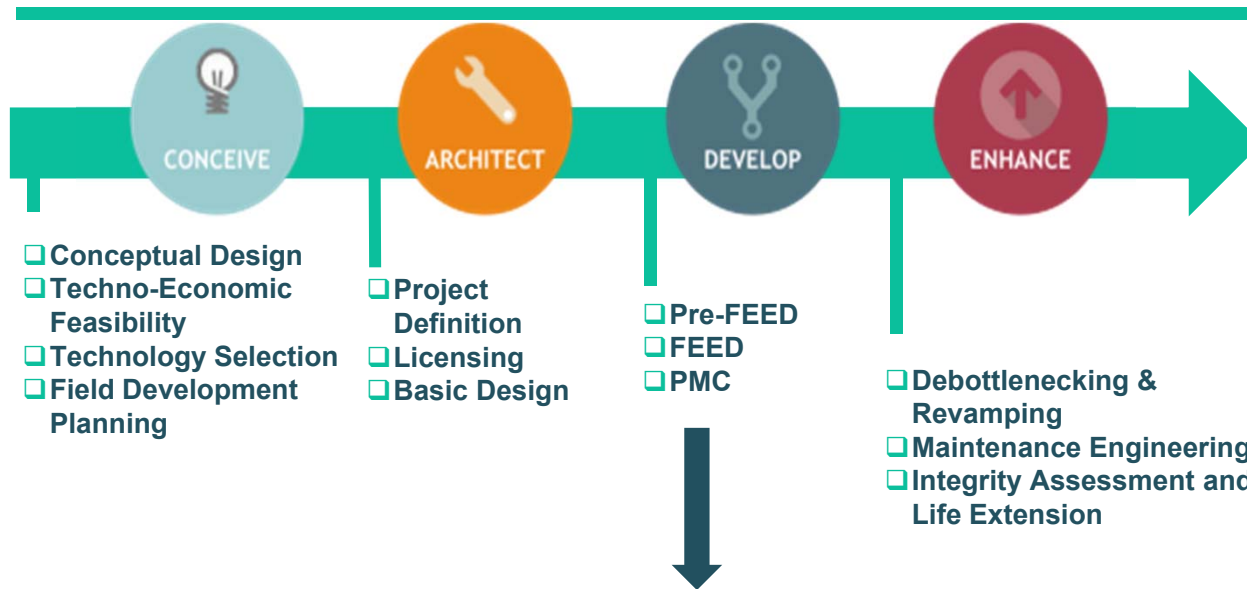
### Customized Products



# Saipem, a global solution provider

## Xsight division

### Engineering Services



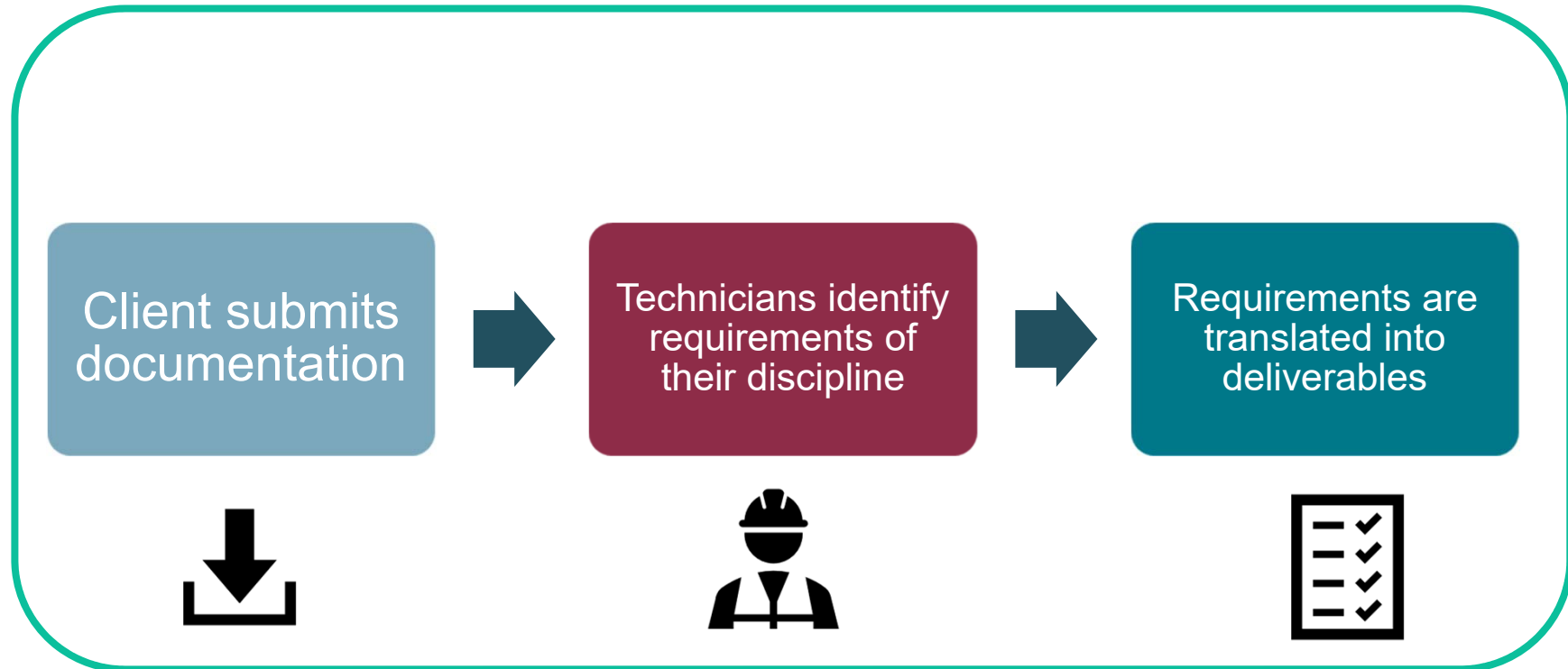
Data & Smart Technology team provides digital solutions to support FEED projects execution

### Customized Products



# FEED project execution

## Identification of project requirements by discipline



# FEED project execution

## Criticalities in the identification of requirements



Projects can consist of a very large number of documents



Requirements search should be as fast as possible



Some requirements are 'hidden' because not explicitated, thus difficult to find



What if we exploit Machine Learning capabilities to automatically identify such requirements?



# FEED project execution

## Our solution to requirements identification

Machine Learning models can be trained on text analytics tasks like text classification

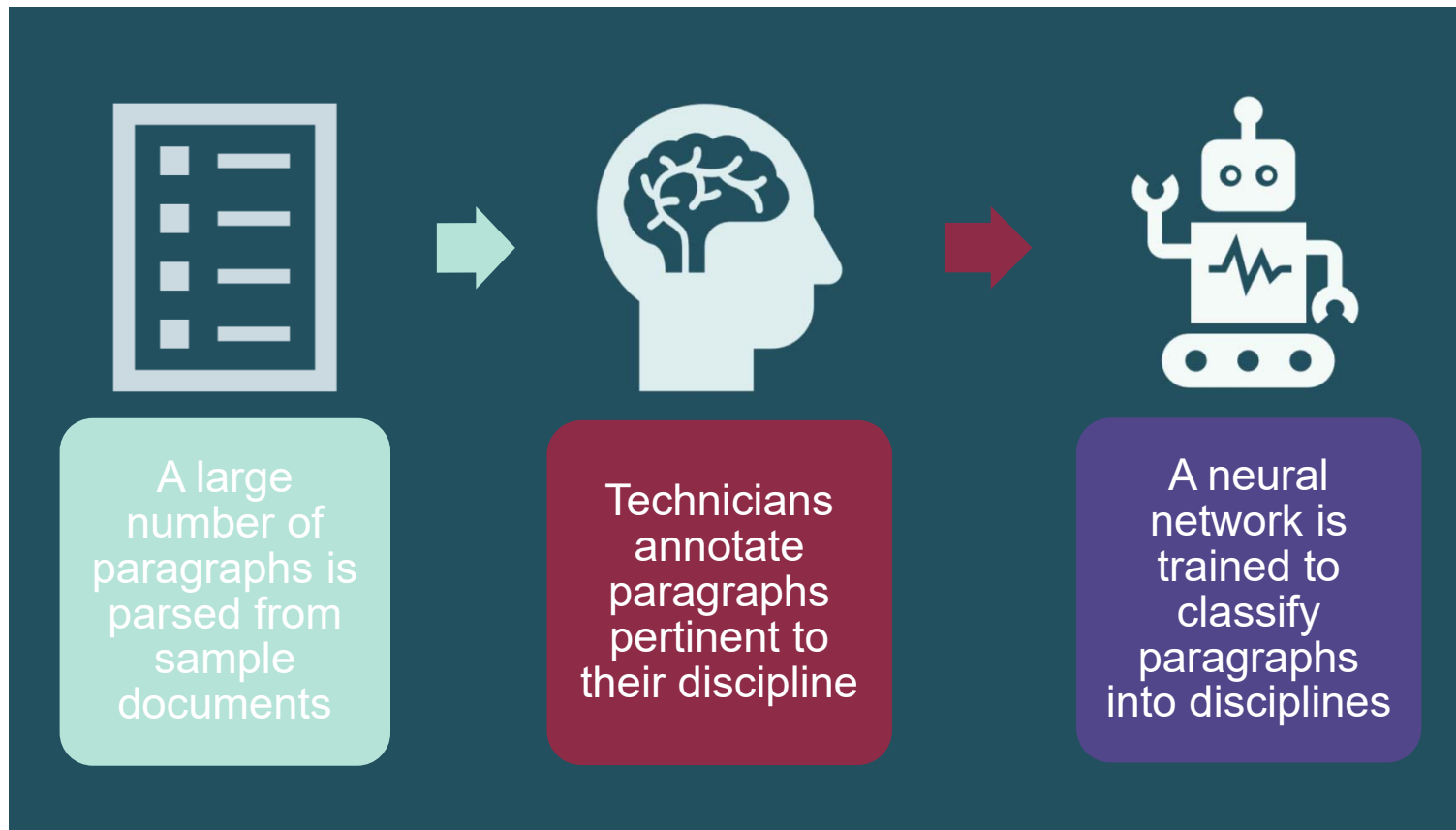
Train a model to classify paragraphs according to the discipline

Each technician has to read all and only the paragraphs pertinent to his discipline



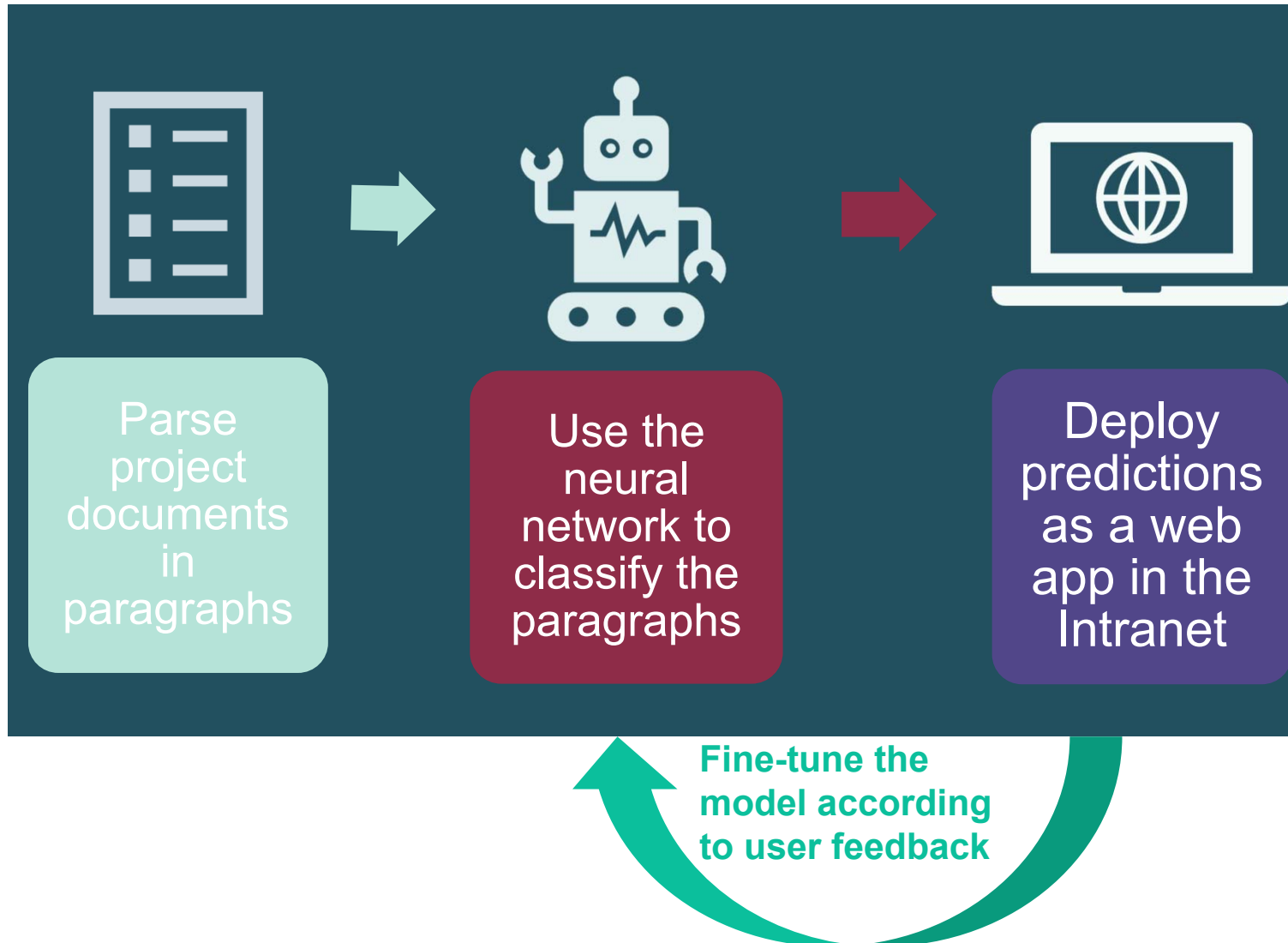
# FEED project execution

## Training the machine learning model



# FEED project execution

## Deployment of the model



# Final considerations and future improvements

All that glitters is not gold...

Having access to such large corpus of text collected over the decades is a huge opportunity and can be considered by itself an example of big data

However, many challenges are hidden in the way of leveraging the data:



Digitalization of paper documents (or low quality scanned PDFs) using OCR technology



Parsing paragraphs from a huge variety of different formattings and page layouts



Non conventional text formats (PPT, Excel, emails, drawings...)



# Thank you

[claudio.nespoli@saipem.com](mailto:claudio.nespoli@saipem.com)

[francesco.ramoni@saipem.com](mailto:francesco.ramoni@saipem.com)



**<SIGHT**

BY SAIPEM