## Introduction

A key part of enabling Ockana’s GenAI solution is digitalising unstructured documents.

We need to process legislation and standards in to clauses and apply metadata to each clause so our users can quickly find the information that is relevant to them.

## Challenge

In the challenge folder you’ll find the Pressure Equipment (Safety) Regulations (PER) in pdf format.

If you go to page 9 you’ll see section 9 (***Design and manufacture in accordance with essential safety requirements***) and section 10 (***Technical documentation and conformity assessment***).

Within these sections there are two clauses e.g. in section 9 there is:

Clause 9.1 - Before placing pressure equipment or an assembly referred to in regulation 6 or 7 on the market or using it for their own purposes, a manufacturer ***must*** ensure that it has been designed and manufactured in accordance with the essential safety requirements.

Clause 9.2 - Before placing pressure equipment or an assembly referred to in regulation 8 on the market or using it for their own purposes, a manufacturer ***must*** ensure that it complies with regulation 8(2).

You’ll see that both these clauses include the word ***must***, this indicates the type of requirement. Must requirements are requirements that are legal requirements for our users. There can also be ***shall*** and ***should*** requirement types.

## Task

Please develop a processor that reads this pdf, parses the clause text and clause numbers, determines the type of requirement and exports to csv.

* You can use any language of your choice.
* Please develop the code with reusability and repeatability in mind.
* If you run out of time in the task please add comments with what you would do next.
* Please return your code in either a zip folder or a link to a github repository with information on how it should be ran.

An example output csv can be found in the challenge folder.