Frequently Asked Questions:

I did not find the corpus in the NER site.
After you registered in the shared task, we will send you a link to download the dataset.

Are we restricted to the task corpus or can we use external corpora?
You cannot use previously labelled dataset or a tagger that is knowledgeable or previously trained to predicate named entities. You can not also use any structured data (from Wikipedia or Wikidata) as this might include NER related information.

Are we restricted (internally) to a particular data format we can use our own format. (In any case, the output format will be as required by the task organisers) ?
We will give the corpus with the IBO format and the output must adhere to the format published in the website(clickable). Internally (inside your training code) you can do anything you like.

Can we extend the team after registration?
Yes, you can add another team member, but firstly, you have to request adding a new member by sending an email and providing us with the details of the new team members, including their names and affiliations. Secondly, you have to submit a formal request for the data again including all the team members.

Can we use pre-trained transformers models (like pretrained Bert) and word representations (like word2vec and ELMo etc.) to perform transfer learning?
Yes, you can use pre-trained transformer models such as BERT and word representations like word2vec and ELMo for transfer learning. In fact, the baseline model we provide is based on BERT.

Can I use linguistics features to enhance the dataset to have part of speech tagging and syntactic layers? (In the code of course)
Regarding the use of linguistic features to enhance the dataset, you are allowed to incorporate part of speech tagging and syntactic layers in your code. However, please note that you cannot use a tagger that is specifically trained or knowledgeable in predicting named entities, either fully or partially.

Is the evaluation code shared? as the IOB tagging is usually evaluated by using a perl utility "conlleval" released by CoNLL.
We utilise the Python package seqeval for evaluation, which offers similar functionality as the conlleval script.

Are there any restrictions on technology and deep learning libraries?
There are no restrictions on the choice of technology or deep learning libraries. You are free to use frameworks of your choice to build your models.