Report:

The report presents two different Language models (LM) for the named entity recognition (NER) task on the English subset of the MultiNERD task, namely Multilingual-BERT-MultiNERD and BERT-medium language.

Experimental-setup:

System	**Classes**	**language-model**
A1	**16**	**Multilingual-BERT-MultiNERD**
A2	**16**	**BERT-medium**
B1	**5**	**Multilingual-BERT-MultiNERD**
B2	**5**	**BERT-medium**
C1	**5**	**A1**
C2	**5**	**A2**

Results:

System	F1-Score	Precision	Recall	Accuracy
A1	0.95097	0.94695	0.95503	0.99019
A2	0.92557	0.92467	0.92647	0.96181
B1	0.96799	0.95980	0.97632	0.99810
B2	0.96211	0.96272	0.96151	0.99031
C1	0.96597	0.95517	0.97701	0.99797
C2	0.96506	0.95941	0.97079	0.99139

Discussion:

System C1 reports the same performance as that of system B1. Similarly, system C2 does not report any significant difference from system B2. This implies that using Systems A1 and A2, which are trained to predict all the 16 classes, does not result in any significant performance gain when used as pre-trained models for systems C1 and C2, which are trained to predict a subset of all the 16 classes. However, evaluating these findings on other datasets and models covering more classes will be beneficial.