Conclusions

• CoNLL#: test set consistent with original annotation guidelines but with significantly less annotation noise
• Please use our corrected test set!
• Models need to improve performance on economy documents and data tables

References

[2] Rethinking Generalization of Neural Models: A Named Entity Recognition Case Study (Fu et al., AAAI 2020)
[3] Identifying Incorrect Labels in the CoNLL-2003 Corpus (Reiss et al., CoNLL 2020)