We are going to use the [roberta-large-ner-english] NER model, because it is the one that fits better with the entities that we are trying to recognize: person_name, organization, location…

https://huggingface.co/Jean-Baptiste/roberta-large-ner-english

We also have been searching for others models, such the Stanford NER, that is done on Java labels sequences of words in a text which are the names of things, such as person and company names, or gene and protein names, but we haven't used this one because of the difficult of the installation.

Another model that we have seen is spaCy that it is a transition-based named entity recognition component. The entity recognizer identifies non-overlapping labelled spans of tokens. We haven't use this model because it doesn't use differ between the differtent types of entities, and you also have to create and specific vocabulary in order to get them ready, if we had more papers to analyze maybe this would be an option.

PAPERS:

Paper 1: A Framework for Characterizing Novel Environment Transformations in General Environments

Results:

Precision: 3/3 all are the same org= 1

Recall: 2/3=0.66

Harmony: : 2*prec*rec/(prec+rec)= 2*1*0.66/(1+0.66)=0.79

As said not a problem of the model.

Paper 2: A Glimpse in ChatGPT Capabilities and its impact for AI research

Results:

Most of the example cited have been done using ChatGPT3.5 MISC , some have use the OpenAi MISC playground using the model text davinci-003. MISC Several parts of the document have been written or corrected using ChatGPT3.5. MISC

</> JSON Output Maximize


```
[
  {
    "entity_group": "MISC",
    "score": 0.9013383388519287,
    "word": " ChatGPT3.5",
    "start": 47,
    "end": 57
  },
  {
    "entity_group": "MISC",
    "score": 0.9900894165039062,
    "word": " OpenAi",
    "start": 77,
    "end": 83
  },
  {
    "entity_group": "MISC",
    "score": 0.9128378629684448,
    "word": "\u0002davinci-003.",
    "start": 115,
    "end": 128
  },
  {
    "entity_group": "MISC",
    "score": 0.8589833974838257,
    "word": " ChatGPT3.5.",
    "start": 196,
    "end": 207
  }
]
```


Precision: ¾ because ChatGpt is treated different=0.75

Recall: This paragraph don't give a lot of info about location or person, so there are many entities that are not included, not a problema of the model = 3/4=0.75

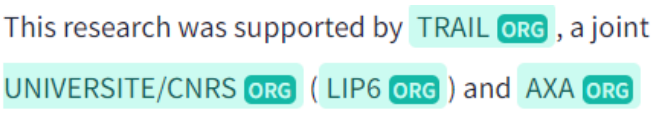
Harmony: : 2*prec*rec/(prec+rec)= 2*0.75*0.75/(0.75+0.75)=0.75

The text doesn't give that much information

Paper 3: Achieving Diversity in Counterfactual Explanations:

a Review and Discussion

Results:



</> JSON Output Maximize


```
[
  {
    "entity_group": "ORG",
    "score": 0.9337700605392456,
    "word": " TRAIL",
    "start": 31,
    "end": 36
  },
  {
    "entity_group": "ORG",
    "score": 0.9950530529022217,
    "word": "\r\nSORBONNE UNIVERSITE/CNRS",
    "start": 64,
    "end": 90
  },
  {
    "entity_group": "ORG",
    "score": 0.9941776394844055,
    "word": "LIP6",
    "start": 92,
    "end": 96
  },
  {
    "entity_group": "ORG",
    "score": 0.9981497526168823,
    "word": " AXA",
    "start": 102,
    "end": 105
  }
]
```


Precision: 4/4=1

Recall: 1.

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 4: A-ePA*SE: Anytime Edge-Based Parallel A* for Slow Evaluations

Results:

This work was supported by the ARL-sponsored A2I2 MISC program, contract W911NF-18-2-0218, and ONR ORG grant N00014-18-1-2775.

</> JSON Output Maximize


```
[
  {
    "entity_group": "MISC",
    "score": 0.9164358377456665,
    "word": " ARL-sponsored A2I2\r\n",
    "start": 31,
    "end": 51
  },
  {
    "entity_group": "ORG",
    "score": 0.9932550191879272,
    "word": " ONR",
    "start": 91,
    "end": 94
  }
]
```


Precision: 3/5=0.6

Recall: 3/5. No mention to authors or the organization=0.6

Harmony: : 2*prec*rec/(prec+rec)=2*0.6*0.6/(0.6+0.33)=0.77

The entities with this numbers are very difficult for the model to be detected

Paper 5: Are ChatGPT and GPT-4 General-Purpose Solvers for Financial Text Analytics? An Examination on Several Typical Tasks

Results:

This research was funded in part by the Faculty Research Awards MISC of J ORG . P. Morgan AI Research. ORG The authors are solely responsible for the contents of the paper and the opinions expressed in this publi□cation do not reflect those of the funding agencies . ORG We also thank Samuel Chan PER for helping prepare this preprint . ORG

</> JSON Output

Maximize


```
[
  {
    "entity_group": "MISC",
    "score": 0.9004910588264465,
    "word": " Faculty\r\nResearch Awards",
    "start": 40,
    "end": 64
  },
  {
    "entity_group": "ORG",
    "score": 0.9922888875007629,
    "word": " J",
    "start": 68,
    "end": 69
  },
  {
    "entity_group": "ORG",
    "score": 0.9964518547058105,
    "word": "P. Morgan AI Research.",
    "start": 70,
    "end": 92
  },
  {
```

```
{
  "entity_group": "ORG",
  "score": 0.9959297776222229,
  "word": ".",
  "start": 252,
  "end": 253
},
{
  "entity_group": "PER",
  "score": 0.9996140599250793,
  "word": " Samuel Chan",
  "start": 269,
  "end": 280
},
{
  "entity_group": "ORG",
  "score": 0.9959297776222229,
  "word": ".",
  "start": 315,
  "end": 316
}
```


Precision: 3/6=0.5

Recall: 3/3. No mention to authors or the organization=1

Harmony: : 2*prec*rec/(prec+rec)=2*0.5*1/(0.5+1)=0.66

Paper 6: Artificial Neuropsychology: Are Large Language ModelsDeveloping Executive Functions?

Results:

I would like to express my gratitude to Pablo Hernan Rodriguez Zivic PER for his guidance, support, and encouragement in the completion of this work, as well as to all the people who participated in the experiment for their time, dedication, and enthusiasm

</> JSON Output Maximize


```
[
  {
    "entity_group": "PER",
    "score": 0.9997878670692444,
    "word": " Pablo Hernan Rodriguez Zivic",
    "start": 40,
    "end": 68
  }
]
```


Precision: 1/1=1

Recall: 1/1=1

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 7: Augmented Large Language Models with Parametric Knowledge Guiding

Results: No acknowledgement paragraph

Paper 8: Building Interoperable Electronic Health Records as Purpose Driven Knowledge Graphs

Results:

The research described in this paper was supported by the InteropEHRate MISC project, a project of the EC ORG Horizon 2020 MISC programme, grant number 826106. We thank all the people from the University of Trento ORG who supported us in the execution of this project, in particular: Danish Asghar Cheema PER, Ronald Chenu Abente PER. The acronym IEHR MISC from the InteropEHRate MISC project, has been freely adapted in this paper as iEHR MISC which stands for interoperable Electronic Health Records MISC.

</> JSON Output Maximize


```
[
  {
    "entity_group": "MISC",
    "score": 0.9866245985031128,
    "word": " InteropEHRate",
    "start": 58,
    "end": 71
  },
  {
    "entity_group": "ORG",
    "score": 0.8076001405715942,
    "word": " EC",
    "start": 99,
    "end": 101
  },
  {
    "entity_group": "MISC",
    "score": 0.9893485903739929,
    "word": " Horizon 2020",
    "start": 102,
    "end": 114
  },
```

```
},
{
  "entity_group": "PER",
  "score": 0.9985764622688293,
  "word": " Danish Asghar Cheema",
  "start": 269,
  "end": 289
},
{
  "entity_group": "PER",
  "score": 0.9986246824264526,
  "word": " Ronald Chenu\r\nAbente",
  "start": 291,
  "end": 311
},
{
  "entity_group": "MISC",
  "score": 0.9461674690246582,
  "word": " IEHR",
  "start": 325,
  "end": 329
},
{
  "entity_group": "MISC",
  "score": 0.9746024012565613,
  "word": " InteropEHRate",
  "start": 339,
  "end": 352
},
{
  "entity_group": "MISC",
  "score": 0.9766431450843811,
  "word": " iEHR",
  "start": 404,
  "end": 408
},
{
  "entity_group": "MISC",
  "score": 0.9455087184906006,
  "word": " Electronic Health\r\nRecords",
  "start": 440,
  "end": 466
```

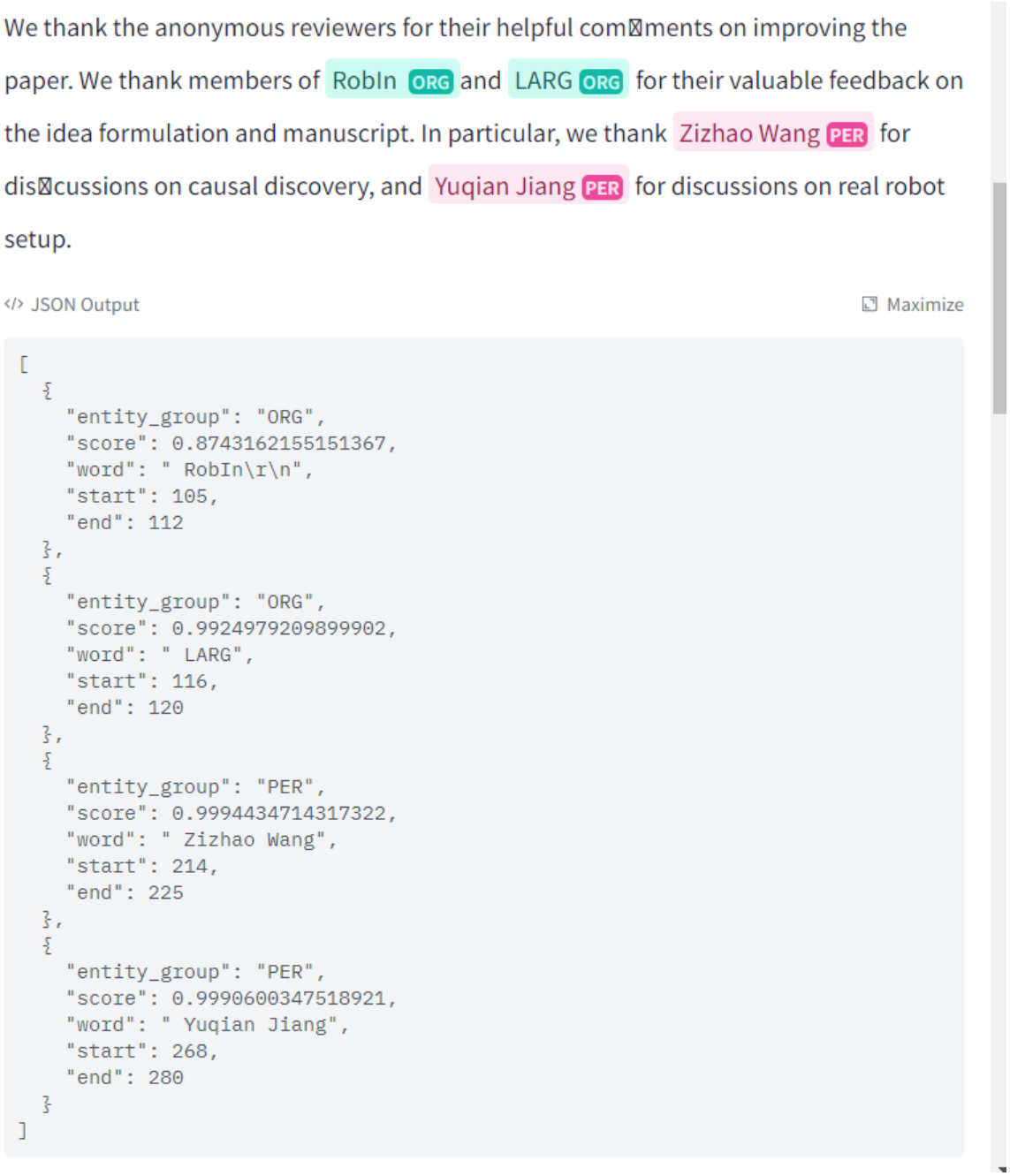

Precision: 9/10=0.9

Recall: 6/8=0.75

Harmony: : 2*prec*rec/(prec+rec)=2*0.9*0.75/(0.9+0.75)=0.81

Paper 9: Causal Policy Gradient for Whole-Body Mobile Manipulation

Results:



Precision: 4/4=1

Recall:4/4=1

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 10: CONTEXT-DEPENDENT COMMUNICATION UNDER ENVIRONMENTAL CONSTRAINTS

Results:

This work was funded by the National Science Centre, Poland ORG (OPUS 15 MISC grant, 2018/29/B/HS1/00884). The ex□periments were partly run on the supercomputers of the Interdisciplinary Centre for Mathematical and Computa□tional Modelling ORG at the University of Warsaw ORG (computa□tional grant G86-1039). The metrics in the experimental results were tracked via Weights & Biases (Biewald PER , 2020)

JSON Output Maximize


```
[
  {
    "entity_group": "ORG",
    "score": 0.9672518968582153,
    "word": " National Science Centre,\r\nPoland",
    "start": 28,
    "end": 60
  },
  {
    "entity_group": "MISC",
    "score": 0.974955141544342,
    "word": "OPUS 15",
    "start": 62,
    "end": 69
  },
  {
    "entity_group": "ORG",
    "score": 0.9911800622940063,
    "word": "\r\nInterdisciplinary Centre for Mathematical and Computa\u0002tior
    "start": 160,
    "end": 232
  },
  {
    "entity_group": "ORG",
    "score": 0.9992039203643799,
    "word": " University of Warsaw",
    "start": 240,
    "end": 260
  },
  {
    "entity_group": "PER",
    "score": 0.9517913460731506,
    "word": "Biewald",
    "start": 370,
    "end": 377
  }
```


Precision: 3/5=0.6

Recall:3/8=0.375

Harmony: : 2*prec*rec/(prec+rec)=2*0.6*0.375/(0.6+0.375)=0.46

Here the Project grants aren't detected

Paper 11: Dual Intent Enhanced Graph Neural Network for Session-based New Item Recommendation

Results:

Precision: 2/2=1

Recall:2/4=0.5

Harmony: : 2*prec*rec/(prec+rec)=2*1*0.5/(1+0.5)=0.66

Don't detect the numbers as miscellaneous entity

Paper 12: EXPLAINING RL DECISIONS WITH TRAJECTORIES

Results:

We thank anonymous reviewers for their helpful feedback to make this work better. Moreover, NJ ORG acknowledges funding support from NSF ORG IIS MISC -2112471 and NSF ORG CARE MISC ER IIS MISC -2141781. Finally, we wish to dedicate this work to the memory of our dear colleague Georgios Theocharous PER who is not with us anymore. While his premature demise has left an unfillable void, his work has made an indelible mark in the domain of reinforcement learning and in the lives of many researchers. He will forever remain in our memories.

</> JSON Output Maximize


```
    "word": " NJ\I\n",
    "start": 92,
    "end": 96
  },
  {
    "entity_group": "ORG",
    "score": 0.8243916630744934,
    "word": " NSF",
    "start": 130,
    "end": 133
  },
  {
    "entity_group": "MISC",
    "score": 0.5714039206504822,
    "word": " IIS",
    "start": 134,
    "end": 137
  },
  {
    "entity_group": "ORG",
    "score": 0.6494945883750916,
    "word": " NSF",
    "start": 150,
    "end": 153
  },
  {
    "entity_group": "MISC",
    "score": 0.5112113356590271,
```

```
 {
   "entity_group": "MISC",
   "score": 0.5714039206504822,
   "word": " IIS",
   "start": 134,
   "end": 137
 },
 {
   "entity_group": "ORG",
   "score": 0.6494945883750916,
   "word": " NSF",
   "start": 150,
   "end": 153
 },
 {
   "entity_group": "MISC",
   "score": 0.5112113356590271,
   "word": " CARE",
   "start": 154,
   "end": 158
 },
 {
   "entity_group": "MISC",
   "score": 0.6143245697021484,
   "word": " IIS",
   "start": 161,
   "end": 164
 },
 {
   "entity_group": "PER",
   "score": 0.9982618689537048,
   "word": " Georgios Theocharous\r",
   "start": 250,
   "end": 271
 }
]
```


Precision: 4/7=0.57

Recall: 4/6=0.66

Harmony: : 2*prec*rec/(prec+rec)=2*0.57*0.66/(0.57+0.66)=0.61

Don’t detect the numbers as a miscellaneous entity

Paper 13: Few-shot Link Prediction on N-ary Facts

Results:

This work is partially supported by the National Natu□ral Science Foundation of China ORG under grants U1911401, 62002341, and 61772501, the GFKJ Innovation Program MISC, and the Lenovo-CAS Joint Lab Youth Scientist Project MISC. The authors would like to thank Zhixuan Li PER, Long Bai PER, and Kailin Zhao PER for their help in writing the paper and drawing the model figure. The authors would also like to thank the anonymous referees for their helpful comments

</> JSON Output Maximize


```
[
  {
    "entity_group": "ORG",
    "score": 0.9989731907844543,
    "word": " National Natu\u0002ral Science Foundation of China",
    "start": 40,
    "end": 85
  },
  {
    "entity_group": "MISC",
    "score": 0.9868348836898804,
    "word": " GFKJ Innovation Program",
    "start": 138,
    "end": 161
  },
  {
    "entity_group": "MISC",
    "score": 0.9862754344940186,
    "word": " Lenovo-CAS Joint Lab Youth Scientist Project",
    "start": 172,
    "end": 216
  },
  {
    "entity_group": "PER",
    "score": 0.9987282752990723,
    "word": " Zhixuan Li",
    "start": 251,
    "end": 261
  },
```


Precision: 6/6=1

Recall:6/9=0.66

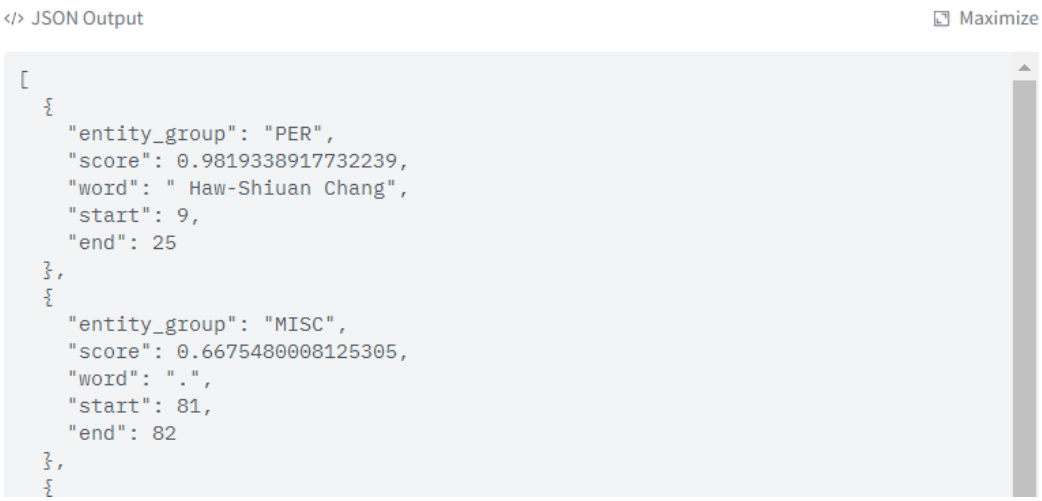
Harmony: : 2*prec*rec/(prec+rec)=2*1*0.5/(1+0.5)=0.79

Don’t detect the numbers as miscellaneous entity

Paper 14: HISTALIGN: Improving Context Dependency in Language Generation by Aligning with History

Results:

We thank Haw-Shiuan Chang PER for helping with pro□viding the ambiguous template data . MISC This work was supported by NSF-CAREER Award MISC 1846185, MISC NSF-AI Engage Institute ORG DRL MISC -2112635, DAR MISC PA ORG Machine Commonsense MISC (MCS MISC) Grant MISC N66001-19- 2-4031, and a Bloomberg Data Science ORG Ph.D. MISC Fel□lowship . MISC The views contained in this article are those of the authors and not of the funding agency



Precision: 4/13=0.3

Recall:4/5=0.8

Harmony: : 2*prec*rec/(prec+rec)=2*0.5*0.8/(0.8+0.5)=0.61

Don't detect the numbers as miscellaneous entity, and sometimes it has problema recognizing the points as entities

Paper 15: How Do In-Context Examples Affect Compositional Generalization?

Results:

We thank all the anonymous reviewers for their valuable comments. This work was supported in part by NSFC ORG under grant No. 62088102.

Precision: 1/1=1

Recall:1/2=0.5

Harmony: : 2*prec*rec/(prec+rec)=2*1*0.5/(1+0.5)=0.66

Don't detect the numbers as miscellaneous entity.

Paper 16: Mixture of personality improved Spiking actor network for efficient multi-agent cooperation

Results:

Precision: 4/4=1

Recall:4/4=1

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 17: Multi-source Education Knowledge Graph Construction and Fusion for College Curricula

Results:

This work is supported by 2022 Beijing Higher Education MISC " Undergraduate Teaching Reform and Innovation Project MISC " and 2022 Education and Teaching Reform Project MISC of Beijing University of Posts and Telecommunications ORG (2022JXYJ-F01).

</> JSON Output Maximize


```
[
  {
    "entity_group": "MISC",
    "score": 0.9608335494995117,
    "word": " 2022 Beijing Higher Education\r",
    "start": 26,
    "end": 56
  },
  {
    "entity_group": "MISC",
    "score": 0.7414656281471252,
    "word": "Undergraduate Teaching Reform and Innovation Project",
    "start": 58,
    "end": 110
  },
  {
    "entity_group": "MISC",
    "score": 0.9814803600311279,
    "word": " 2022 Education and Teaching Reform Project",
    "start": 117,
    "end": 159
  },
  {
```


Precision: 4/5=0.8

Recall:4/4=1

Harmony: : 2*prec*rec/(prec+rec)=2*0.8*1/(0.8+1)=0.88

Problem with separating an entity in two

Paper 18: Patchwork Learning: A Paradigm Towards Integrative Analysis across Diverse Biomedical Data Sources

Results:

This work is supported by NSF ORG awards with number 1750326, 2212175, and NIH ORG awards with number RF1AG072449, R01 MH MISC 124740 and R01 AG MISC 080991.

</> JSON Output Maximize


```
[
  {
    "entity_group": "ORG",
    "score": 0.9919240474700928,
    "word": " NSF",
    "start": 26,
    "end": 29
  },
  {
    "entity_group": "ORG",
    "score": 0.9970831274986267,
    "word": " NIH",
    "start": 71,
    "end": 74
  },
  {
    "entity_group": "MISC",
    "score": 0.5937716960906982,
```


Precision: 2/4=0.5

Recall:2/7=0.28

Harmony: : 2*prec*rec/(prec+rec)=2*0.5*0.28/(0.5+0.28)=0.35

Again problems with numbers as a miscellanous entity.

Paper 19: Predicting the Price Movement of Cryptocurrencies Using Linear Law-based TransformationResults:

Project no . ORG PD MISC 142593 was implemented with the support provided by the Ministry of Cul⊠ture and Innovation of Hungary ORG from the National Research, Development ORG , and Innovation Fund ORG , financed under the PD 22 MISC " OTKA MISC " funding scheme . A.J. ORG received support from the Hungarian Scientific Research Fund ORG (OTKA/NRDI Office ORG) under contract number K123815 . ORG The research 7 was supported by the Ministry of Innovation and Technology NRDI Office ORG within the framework of the MILAB Artificial Intelligence National Laboratory Program MISC . ORG

Precision: 7/14=0.5

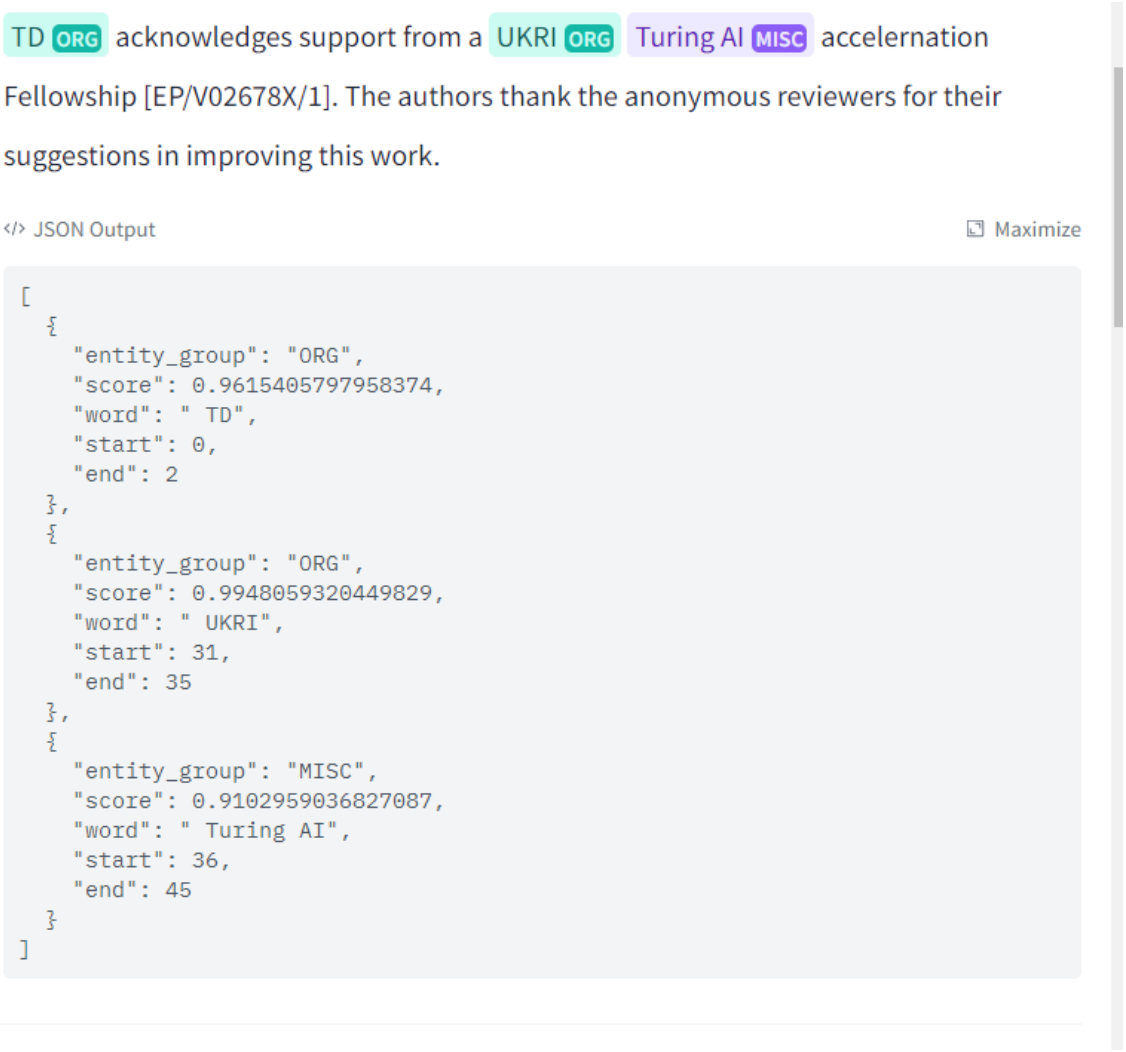
Recall:7/11=0.63

Harmony: : 2*prec*rec/(prec+rec)=2*0.57*0.72/(0.57+0.72)=0.55

Again problems with numbers as a miscellanous entity, and puntual problems with the points org.

Paper 20: Quantifying Consistency and Information Loss for Causal Abstraction Learning

Results:



Precision: 2/3=0.66

Recall:2/2=1

Harmony: : 2*prec*rec/(prec+rec)=2*0.66*1/(0.66+1)=0.79

Problems recognizing some entities

Paper 21: RECKONING: Reasoning through Dynamic Knowledge Encoding

Results:

Precision: 7/7=1

Recall:7/8=0.875

Harmony: : 2*prec*rec/(prec+rec)=2*1*0.875/(1+0.875)=0.93

Problem with the proyect PFFS

Paper 22: Scan2LoD3: Reconstructing semantic 3D building models at LoD3 using ray casting and Bayesian networks

Results:

This work was supported by the Bavaria MISC n State Ministry for Economic Affairs, Regional Development and Energ ORG y within the framework of the IuK Bayer ORG n project MoFa3 MISC D - Mobile Erfassung von Fassaden mittels 3D Punktwolken, Grant No. IUK643/001. Moreover, the work was conducted within the framework of the Leonhard Obermeyer Cente ORG r at the Technical University of Munic ORG h (TU ORG M).

Precision: 6/7*=0.85

Recall:6/8=0.75

Harmony: : 2*prec*rec/(prec+rec)=2*0.85*0.75/(0.85+0.75)=0.79

*It looks that with this paragraph have some bug regarding the last letter I won't be counting as a mistake, because is the only time I have seen it

Paper 23: Score: A Rule Engine for the Scone Knowledge Base System

Results:

I would like to thank Professor Scott Fahlman PER for introducing me to Scone ORG and providing me with guidance on how a rule engine in Scone LOC should be designed. His insights on the overall structure of Scone ORG as well as on other related work in production systems were instrumental in helping me write this thesis. I would also like to thank Alessandro Oltramari PER for agreeing to take the time to be on the thesis committee and for making corrections and commenting on areas of improvement for this document. Lastly, I would like to thank my sister, my parents, and my friends for always giving me their support when I needed it

Precision: 4/5=0.8

Recall:3/4=0.75

Harmony: : 2*prec*rec/(prec+rec)=2*0.8*0.75/(0.8+0.75)=0.77

Paper 24: Sketching the Future (STF): Applying Conditional Control Techniques to Text-to-Video Models

Results:

We would like to thank Professor Pathak PER and the course staff of Visual Learning and Recognition ORG for their support, and Mrinal Verghese PER for his compute resources. Also we would like to thank ChatGPT ORG for assisting with the writing and organization of this paper.

Precision: 4/4=1

Recall:4/4=1

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 25: Stackelberg Games for Learning Emergent Behaviors During Competitive Autocurricula

Results: No acknowledgement paragraph

Paper 26: S T A R C O D E R : M AY T H E S O U R C E B E W I T H Y O U !

Results:

We would like to thank HuggingFace ORG for providing the compute resources to train the StarCoder MISC models. We also thank Suriya Gunasekar PER for help with the data inspection, and Sebastien Paquet PER for proofreading this work.

</> JSON Output Maximize


```
[
  {
    "entity_group": "ORG",
    "score": 0.9779202938079834,
    "word": " HuggingFace",
    "start": 23,
    "end": 34
  },
  {
    "entity_group": "MISC",
    "score": 0.9957900047302246,
```


Precision: 4/4=1

Recall:4/4=1

Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 27: The Current State of Summarization

Results:

Precision: 1/1=1

Recall:1/1=1

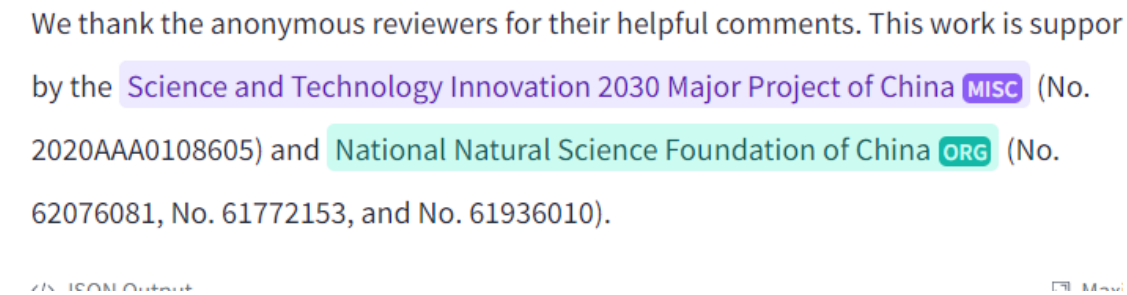
Harmony: : 2*prec*rec/(prec+rec)=2*1*1/(1+1)=1

Paper 28: Train a Real-world Local Path Planner in One Hour via Partially Decoupled Reinforcement Learning and Vectorized Diversity

Results: No acknowledgement paragraph

Paper 29: U-NEED: A Fine-grained Dataset for User Needs-Centric E-commerce Conversational Recommendation

Results:



Precision: 2/2=1

Recall:2/5=0.4

Harmony: : 2*prec*rec/(prec+rec)=2*1*0.4/(1+0.4)=1

Again some problems creating an entity with numbers

Paper 30: Wasserstein-Fisher-Rao Embedding: Logical Query Embeddings with Local Comparison and Global Transport

Results:

The authors of this paper were supported by the NSFC Fund MISC (U20B2053) from the NSFC ORG of China LOC, the RIF MISC (R6020-19 and R6021-20) and the GRF MISC (16211520 and 16205322) from RGC ORG of Hong Kong LOC, the MHKJFS ORG (MHP/001/19) from ITC ORG of Hong Kong LOC and the National Key R&D Program of China MISC (2019YFE0198200) with spe□cial thanks to HKMAAC ORG and CUSBLT ORG. We also thank the support from NVIDIA AI Technology Center ORG (NVAITC ORG) and the UGC ORG Research Match□ing Grants MISC (RMGS MISC 20EG01-D, RMGS MISC 20CR11, RMGS MISC 20CR12, RMGS MISC 20EG19, RMGS MISC 20EG21, RMGS MISC 23CR05, RMGS MISC 23EG08)

It has the same problema as the other but in a big way, not gonna take in account

Conclusions:

With all the data taken we have listed some pros and cons of our model:

Pros:

- High Accuracy: Roberta-Large-NER-English has been pre-trained on a large corpus of English text data, which has enabled it to achieve high accuracy in identifying named entities in text.
- Generalizability: The pre-training process of Roberta-Large-NER-English has made it capable of identifying named entities in a wide range of text genres, such as news articles, academic papers, social media posts, etc.
- Customizable: The model can be fine-tuned on a specific domain or task to further improve its accuracy and suitability for that particular use case.

Cons:

- Computationally Expensive: The model is large and computationally expensive, which means that it may require a powerful computing infrastructure to run effectively.
- Requires Large Amounts of Training Data: The pre-training process of the model requires a large amount of text data, which may not always be readily available for specific domains or use cases.
- May Struggle with Domain-Specific Terminology: Since the model is pre-trained on general English text data, it may struggle to identify named entities that are specific to a particular domain or industry.