**Motivation**

- LMs have rich entity knowledge about their training data, but as new entities emerge, this knowledge becomes outdated. **How can we test what models know about new entities?**
- Existing benchmarks usually test KB triples. **How can we test a broader set of inferences about entities?**

**Data Collection**

- **Date Indexing**
  ECBD is a cloze-style task. Each masked span is associated with an entity which is paired with its origination year.

- **Diverse Masked Spans**
  NP spans primarily represent relational knowledge. Random spans cover other types of knowledge (e.g., actions an entity can take).

**What do large LMs know about new entities that appeared after pre-training?**

The **[Windows 11]** name was accidentally released in **[MASK]** in June 2021. NP: an official Microsoft support document

The **[GPT-2]** model has **[MASK]**, and was trained on 8 million web pages. Entity from 2019 NP: 1.5 billion parameters

**[Austin FC]** are the operators of a newly-**[MASK]**. Entity from 2018 Random: built stadium at McKalla Place

**[Hurricane Irma]** affected multiple states in the **[MASK]**. Entity from 2017 Random: South, especially Florida

**Data Statistics**

<table>
<thead>
<tr>
<th>Year</th>
<th># Sent.</th>
<th># Ent.</th>
<th>Span Len.</th>
<th>[Span V.]</th>
</tr>
</thead>
<tbody>
<tr>
<td>2017</td>
<td>35k</td>
<td>2,106</td>
<td>2.9</td>
<td>19,542</td>
</tr>
<tr>
<td>2018</td>
<td>8k</td>
<td>1,910</td>
<td>2.9</td>
<td>8,644</td>
</tr>
</tbody>
</table>

* Set of existing entities to use a baseline.

**New Entity Types**

- Misc. 4.0%
- Other Events 11.0%
- Organizations 15.0%
- Businesses 10.0%
- Natural Risks 7.0%
- Sports 19.0%
- Media 21.0%
- Infrastructure 9.0%

**Experiments**

- **What do LMs know about seen/unseen entities?**
  - Compare T5-Large with different inputs.
    - No Ent: remove entity mention, Definition: add definition from EN Wikipedia
    - Measure perplexity on the masked spans

- **Observation**
  - Newer entities may be harder.
  - Adding definition is particularly helpful for new entities – Entity knowledge matters for this task!

Data is available at github.com/yasumasaonoe/ecbd