Entity Cloze By Date: What LMs Know About Unseen Entities
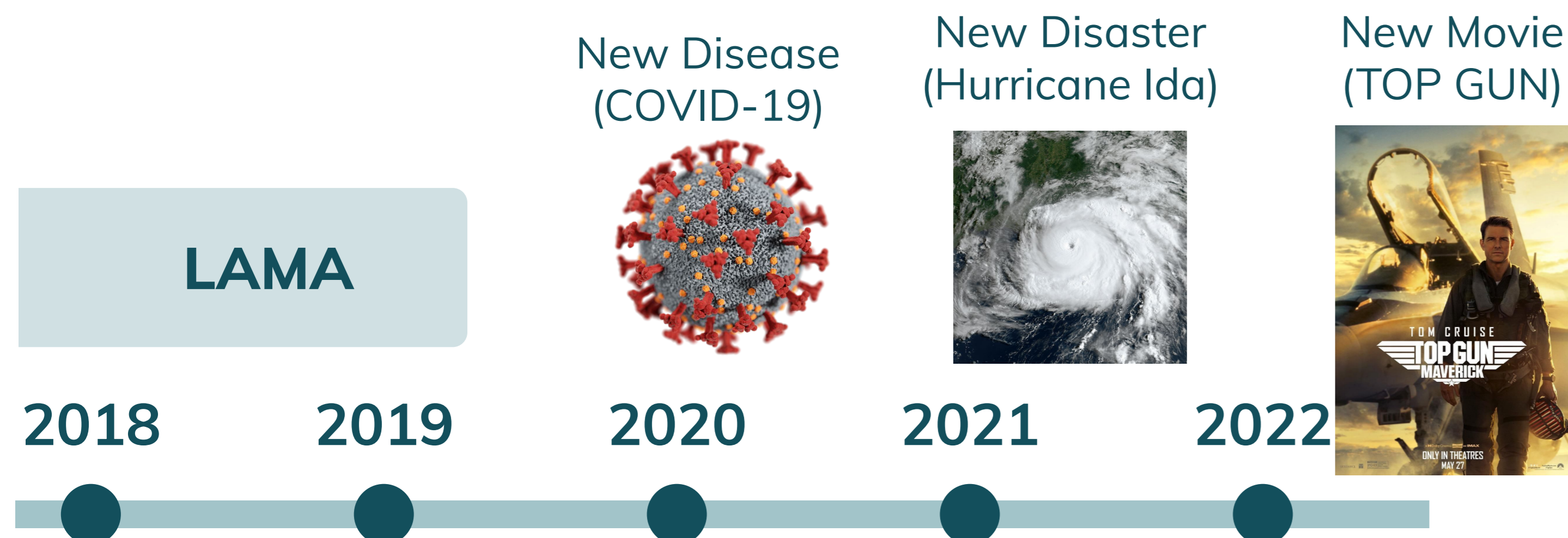
Yasumasa Onoe, Michael J.Q. Zhang, Eunsol Choi, Greg Durrett
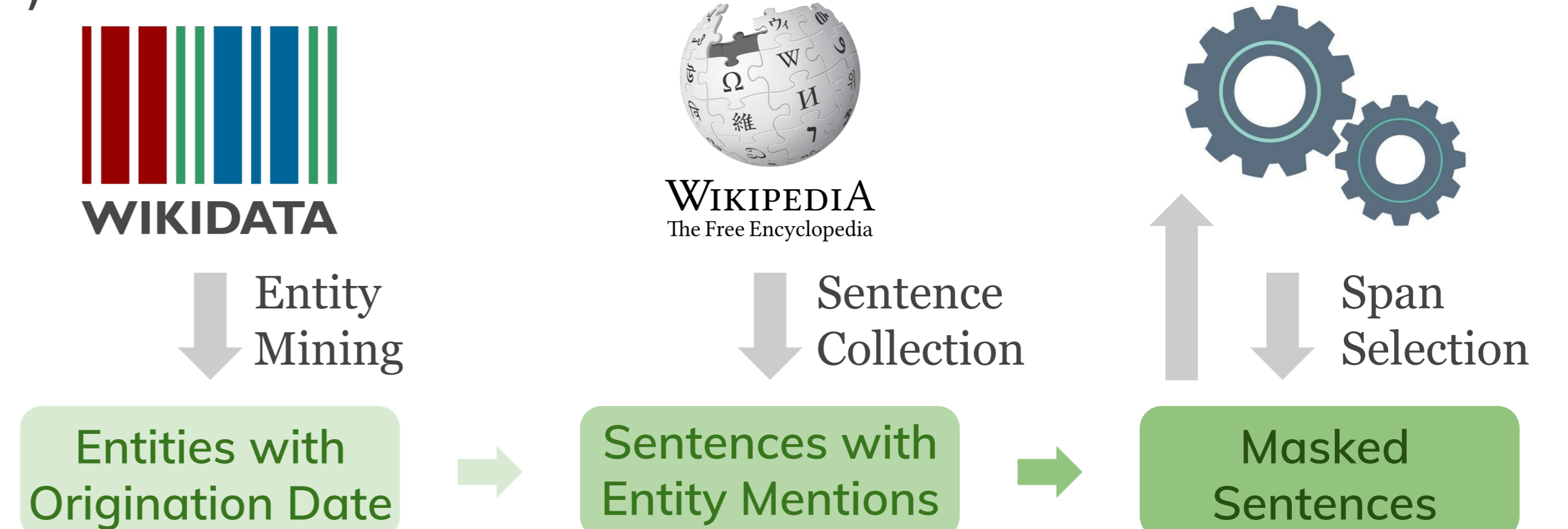
TEXAS — The University of Texas at Austin

TAUR LAB

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- LMs have rich entity knowledge about their training data, but as new entities emerge, this knowledge becomes outdated.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?**

New Disease (COVID-19) — New Disaster (Hurricane Ida) — New Movie (TOP GUN)

LAMA

2018 2019 2020 2021 2022

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).

WIKIDATA → Entity Mining → Entities with Origination Date

WIKIPEDIA The Free Encyclopedia → Sentence Collection → Sentences with Entity Mentions

Span Selection → Masked Sentences

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.
Entity from 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**

ECBD ECBD ECBD ECBD ECBD

2017 2018 2019 2020 2021 2022

BERT
GPT2
T5
GPT3
GPT-Neo
PaLM

- **Task**
Given a cloze sentence with **a new entity,** predict **masked tokens** (measure the model's perplexity)

- **Seen/Unseen Entities**
Test examples are grouped by the origination year. It's easy to test LMs on **OLD/NEW** entities.

Automatically constructed; we can update ECBD easily!

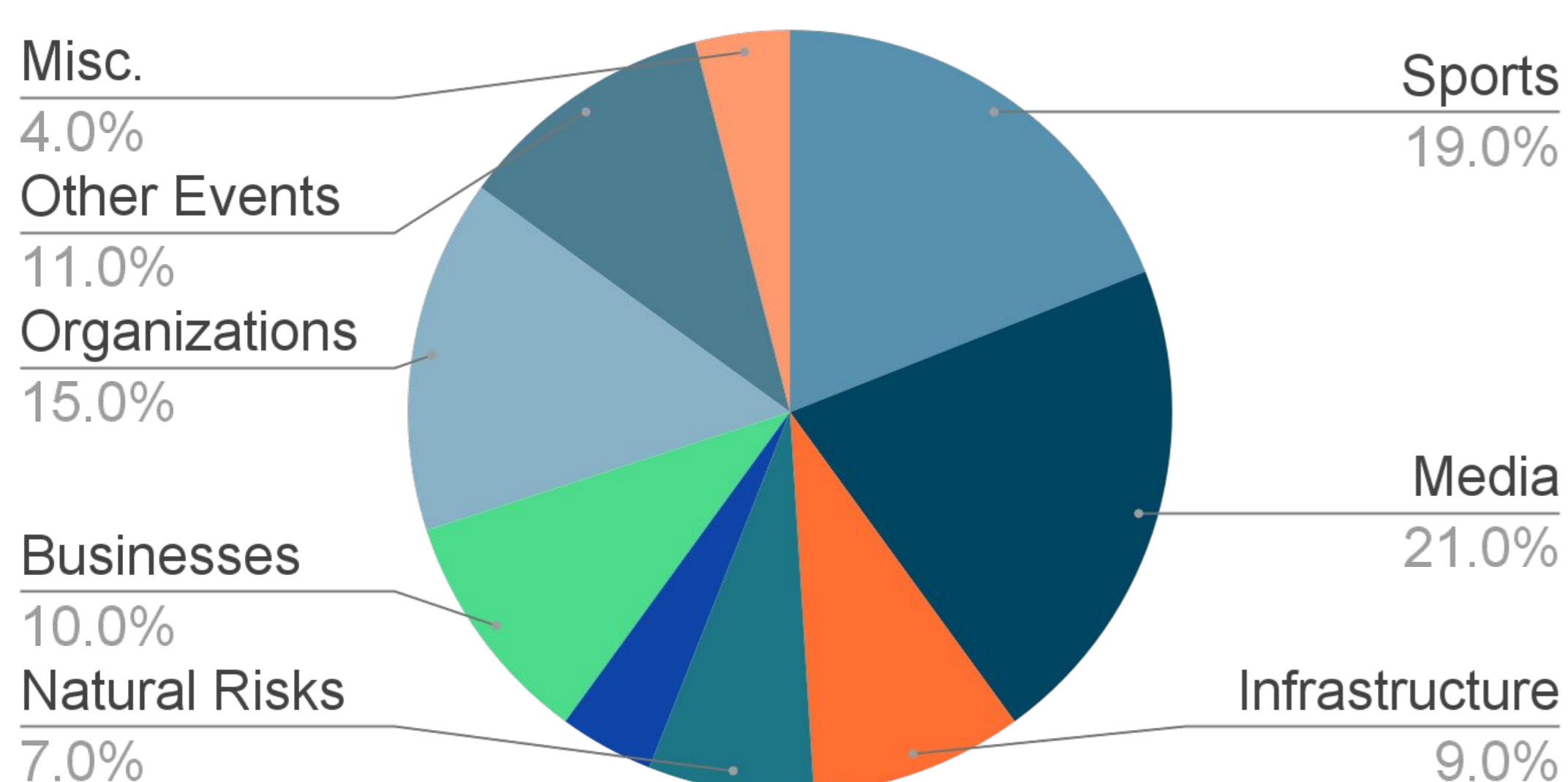**Can test if knowledge editing approaches successfully inject information**
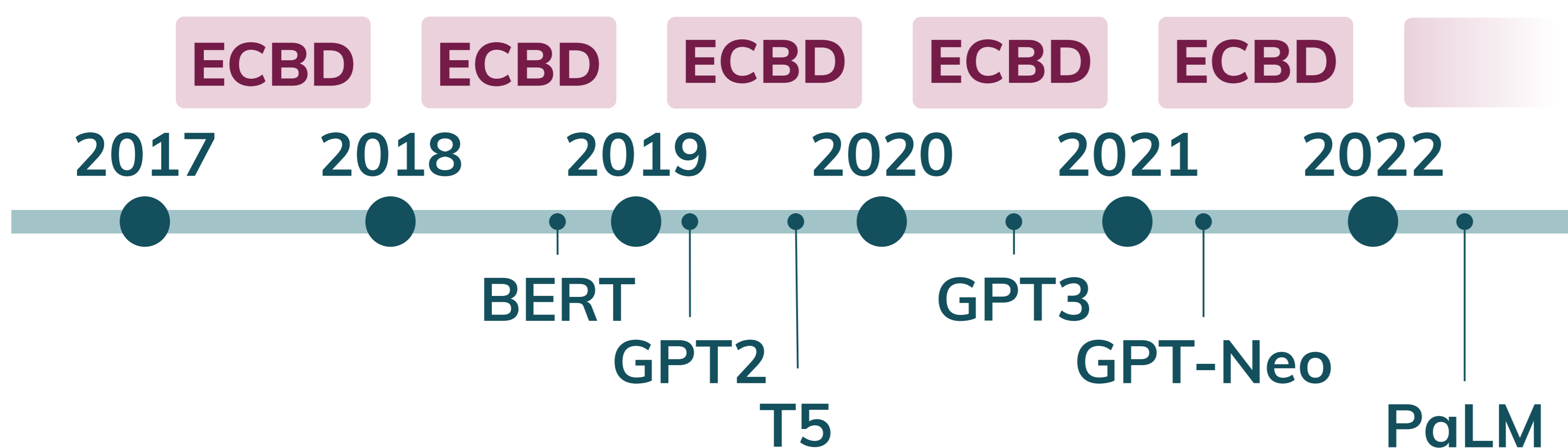
Data Statistics

- **Size**

	# Sent.	# Ent.	Span Len.	|Span V.|
ECBD	35k	2,106	2.9	19,542
Popular*	8k	1,910	2.9	8,644

* 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!**

Legend: No Ent. | Original | Random Def. | Definition

Perplexity

Popular: 18.28 | 13.02 | 12.1 | 11.04
2017-2019: 22.35 | 15.39 | 14.33 | 11.73
2020-2021: 26.69 | 19.43 | 17.34 | 13.6