GoLLIE: Annotation Guidelines Improve Zero-Shot Information-Extraction

Oscar Sainz, Iker García-Ferrero, Rodrigo Agerri, Oier Lopez de Lacalle, German Rigau, Eneko Agirre

ixa — Universidad del País Vasco / Euskal Herriko Unibertsitatea — HiTZ Hizkuntza Teknologiako Zentroa / Basque Center for Language Technology

WHY GUIDELINES?

LLMs are being used to extract information using the label itself (e.g. Person). But many extraction tasks are nuanced, and guidelines define what the user is interested in. For instance, extract persons in this sentence, where persons are defined as in CoNLL03 *"First, middle and last names of people or animals"*.

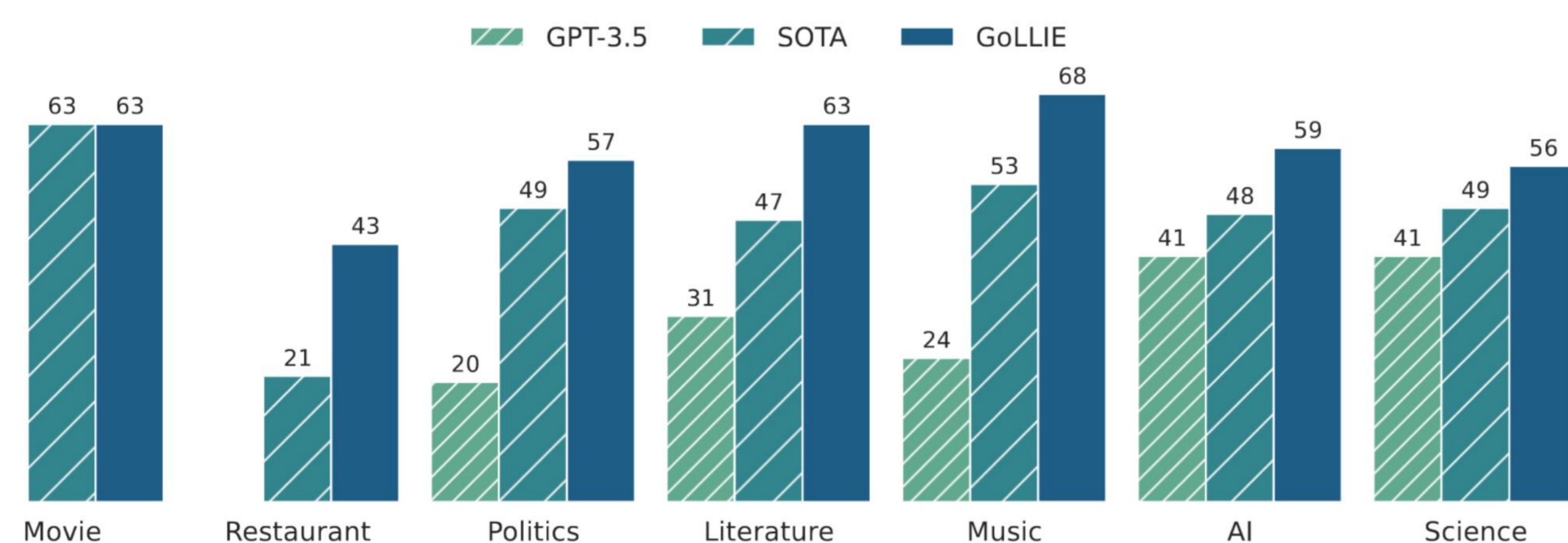The name of [TED]$_{PER}$'s dog is [Luna]$_{PER}$.

A LLM which cannot process properly the guidelines would fail to extract Luna, contrary to the guidelines.

PROPOSED APPROACH

- Code format — Python— to represent the annotation schema as classes. Already familiar for both LLMs and humans!
- Introduce detailed annotation guidelines into the schema representation.
- Train the model to follow the guidelines by several proposed regularization techniques.

HOW DOES GoLLIE PERFORM?

- GoLLIE outperforms GPT-3.5 on zero-shot domain-specific benchmarks - both using guidelines.
- GoLLIE improves over SOTA models that do not use any guidelines.

GPT-3.5 / SOTA / GoLLIE

	Movie	Restaurant	Politics	Literature	Music	AI	Science
GPT-3.5	63	21	20	31	24	41	41
SOTA	63	43	49	47	53	48	49
GoLLIE	—	—	57	63	68	59	56

https://github.com/hitz-zentoa/GoLLIE

https://hf.co/HiTZ/GoLLIE-{7, 13, 34}B

INPUT-OUTPUT REPRESENTATION

Class name masking — *Launcher -> Label_1*

Guideline paraphrasing — *Generate paraphrases using LLMs*

Class order shuffling — *Shuffle classes for each example*

Class dropout — *Remove random classes (input and output)*

**Candidate Sampling** — *"Venus", "High-orbit", "Uranus"*

```python
@dataclass
class Launcher(Template):
    """Refers to a vehicle designed primarily to transport payloads from
    the Earth's surface to space. Launchers can carry various payloads,
    including satellites, crewed spacecraft, and cargo. They are usually
    multi-stage vehicles that use rocket engines for propulsion."""

    mention: str # The name of the launcher vehicle. Such as: "Sturn V",
    "Atlas V", "Soyuz", "Ariane 5"
    space_company: Optional[Organization] = None # The company that
    operates the launcher. Such as: "Blue origin", "ESA", "Boeing"
    crew: Optional[List[Person]] = None # Names of the crew members
    boarding the Launcher. Such as: "Neil Armstrong", "Michael Collins",
    ... rin"

class Mission(Template):
    """Any planned or accomplished journey beyond Earth's atmosphere with
    specific objectives, either crewed or uncrewed. It includes missions
    to satellites, the International Space Station (ISS), other celestial
    bodies, and deep space."""

    mention: str # The name of the mission. Such as: "Apollo 11",
    "Artemis", "Mercury"
    date: Optional[str] = None # The start date of the mission
    destination: Optional[Location] = None # The place or planet to which
    the launcher will be sent. Such as "Moon", "low-orbit", "Saturn"

# This is the text to analyze
text = "The Ares 3 mission to Mars is scheduled for 2032. The Starship
rocket build by SpaceX will take off from Boca Chica, carrying the
astronauts Max Rutherford and Elena Soto"

----------------------- (↑Input, ↓Output) -----------------------

result = [
    Launcher("Starship",
        space_company=Organization("SpaceX"),
        crew=[Person("Max Rutherford"), Person("Elena Soto")],
    ),
    Mission("Ares 3",
        date="2032",
        destination=Location("Mars"),
    ),
]
```

EVALUATION SETUP

Dataset	Domain	NER	RE	EE	EAE	SF	Training	Evaluation
ACE05 (Walker et al., 2006)	News	✓	✓	✓	✓		✓	✓
BC5CDR (Wei et al., 2016)	Biomedical	✓					✓	✓
CoNLL 2003 (Tjong Kim Sang & De Meulder, 2003)	News	✓					✓	✓
DIANN (Fabregat et al., 2018)	Biomedical	✓					✓	✓
NCBIDisease (Islamaj Doğan & Lu, 2012)	Biomedical	✓					✓	✓
Ontonotes 5 (Pradhan et al., 2013)	News	✓					✓	✓
RAMS (Ebner et al., 2020)	News				✓		✓	✓
TACRED (Zhang et al., 2017)	News					✓	✓	✓
WNUT 2017 (Derczynski et al., 2017)	News	✓					✓	✓
BroadTwitter (Derczynski et al., 2016)	Twitter	✓						✓
CASIE (Satyapanich et al., 2020)	Cybercrime			✓	✓			✓
CrossNER (Liu et al., 2021b)	*Many*	✓						✓
E3C (Magnini et al., 2021)	Biomedical	✓						✓
FabNER (Kumar & Starly, 2022)	Science	✓						✓
HarveyNER (Chen et al., 2022)	Twitter	✓						✓
MIT Movie (Liu et al., 2013)	Queries	✓						✓
MIT Restaurants (Liu et al., 2013)	Queries	✓						✓
MultiNERD (Tedeschi & Navigli, 2022)	Wikipedia	✓						✓
WikiEvents(Li et al., 2021)	Wikipedia	✓		✓	✓			✓

SEEN VS UNSEEN LABELS

We compare GoLLIE against not using guidelines.

Major improvements when evaluating on unseen labels.

baseline / 7B / 13B / 34B
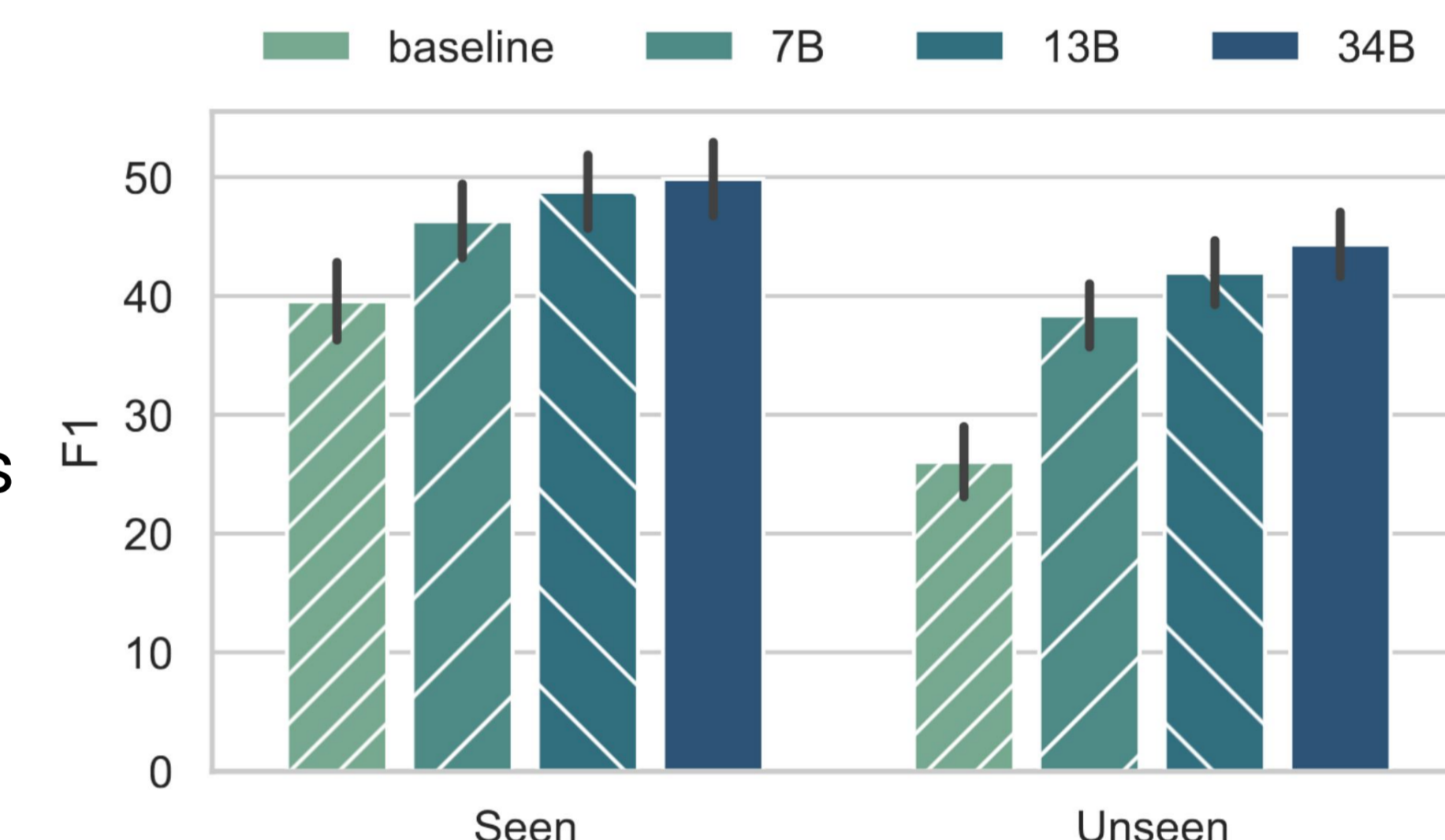
ERROR ANALISIS

- Green: improve the results.
- Blue: makes no difference.
- Red: Adding guidelines does not solve the underlying problem.

Dataset	Label	Guideline	Baseline	🐶
MultiNERD	Media	Titles of films, books, songs, albums, fictional characters and languages.	13.6	69.1
CASIE	Vul. Patch	When a software company addresses a vulnerability by releasing an update.	27.7	70.5
Movie	Trailer	Refers to a short promotional video or preview of a movie.	00.0	76.4
AI	Task	Particular research task or problem within a specific AI research field.	02.7	63.9
MultiNERD	Time	Specific and well-defined time intervals, such as eras, historical periods, centuries, years and important days.	01.4	03.5
Movie	Plot	Recurring concept, event, or motif that plays a significant role in the development of a movie.	00.4	05.1
AI	Misc	Named entities that are not included in any other category.	01.1	05.2
Literature	Misc	Named entities that are not included in any other category.	03.7	30.8
Literature	Writer	Individual actively engaged in the creation of literary works.	04.2	65.1
Literature	Person	Person name that is not a writer.	33.5	49.4
Science	Scientist	A person who is studying or has expert knowledge of a natural science field.	02.1	05.8
Science	Person	Person name that is not a scientist.	46.1	45.9
Politics	Polit. Party	Organization that compete in a particular country's elections.	11.2	34.9