

factgenie:

A Framework for Span-based Evaluation of Generated Texts

Zdeněk Kasner

Ondřej Plátek

Patrícia Schmidtová

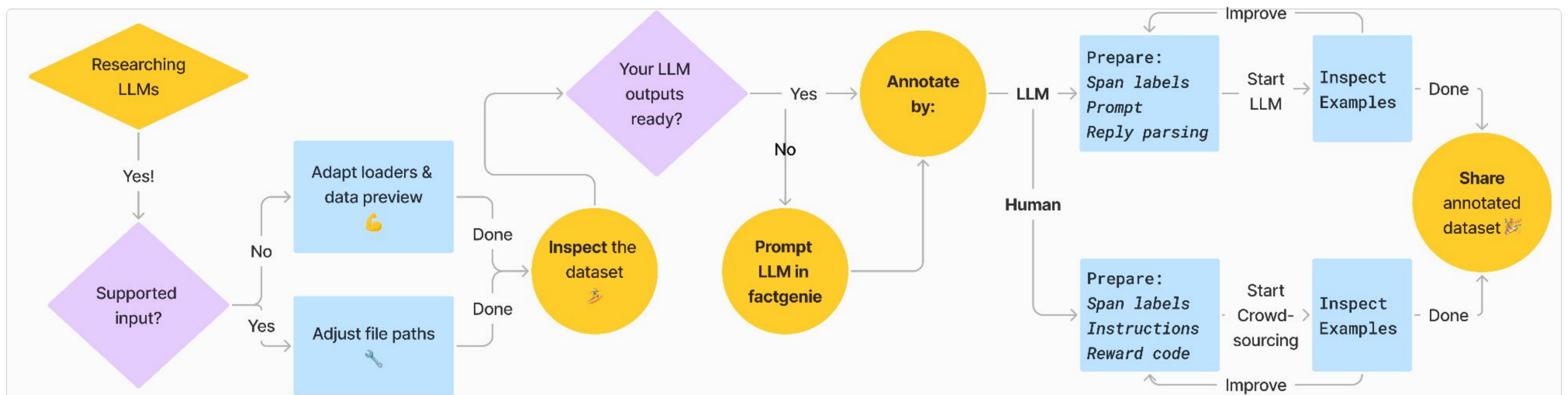
Simone Balloccu

Ondřej Dušek

{platek,kasner}@ufal.mff.cuni.cz

Evaluate LLMs interactively, effortlessly with crowd workers or LLMs

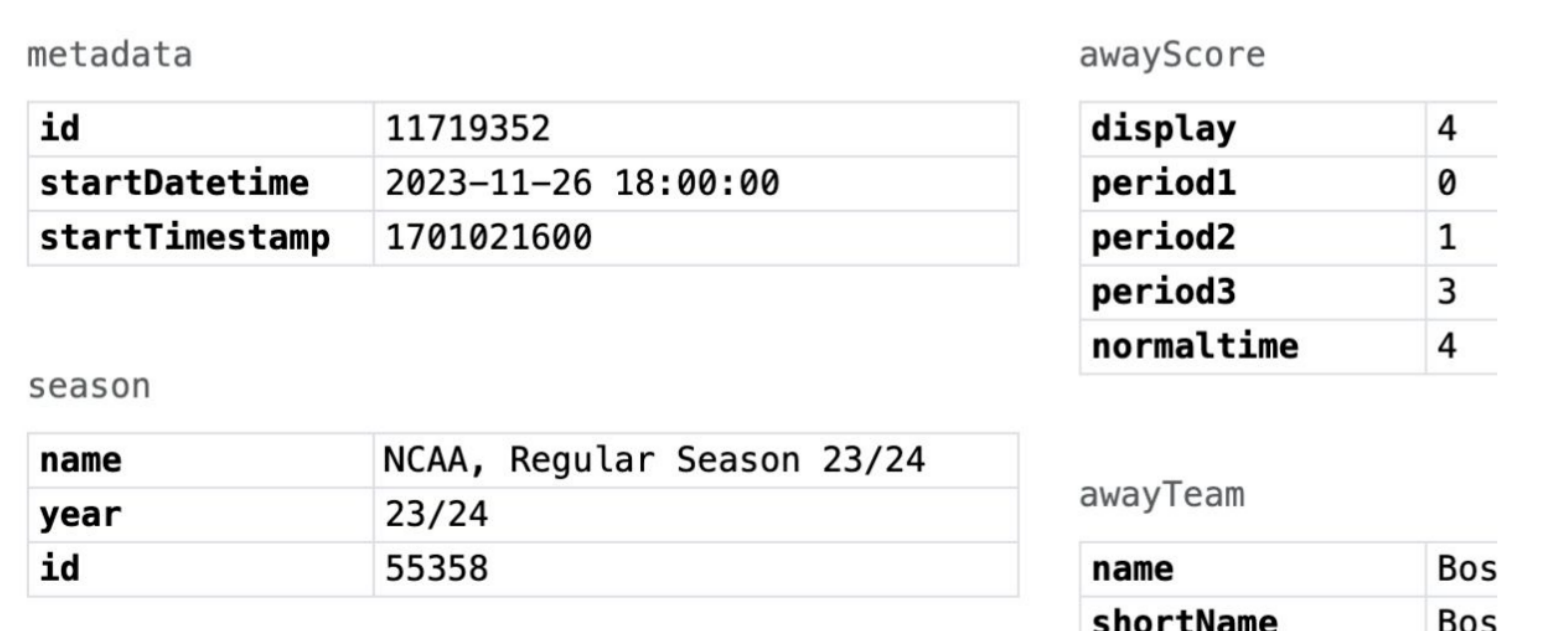
Context and errors matter! Use factgenie to visualize them clearly!



Demo Datasets

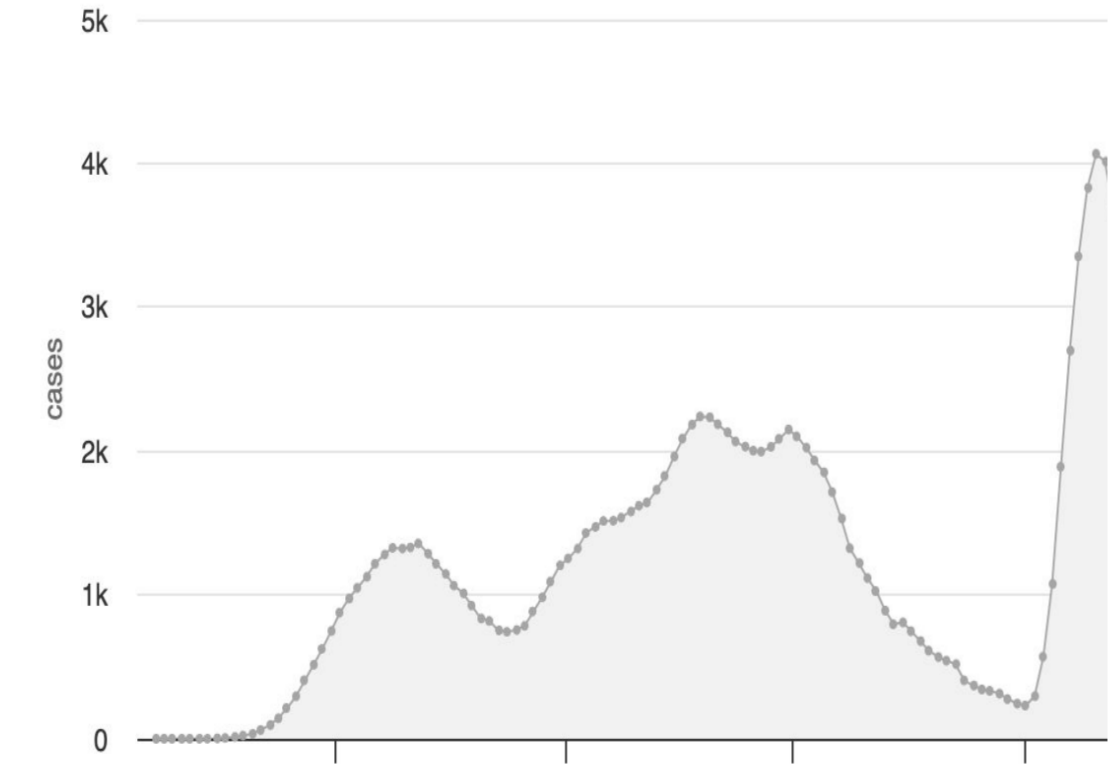
Dataset	Loader	Task	Hard to add?
xsum	PlainText	summarization	easy!
easy VQA	HTML	visual QA	OK
openweather	Custom	graph-to-text	OK
gsmarena	Custom	table-to-text	easy!
wikidata	Custom	rdf-to-text	easy!
logicnlg	Custom	table-to-text	OK
yours!	is	to be added	easily!

Harvard University – Boston College Eagles

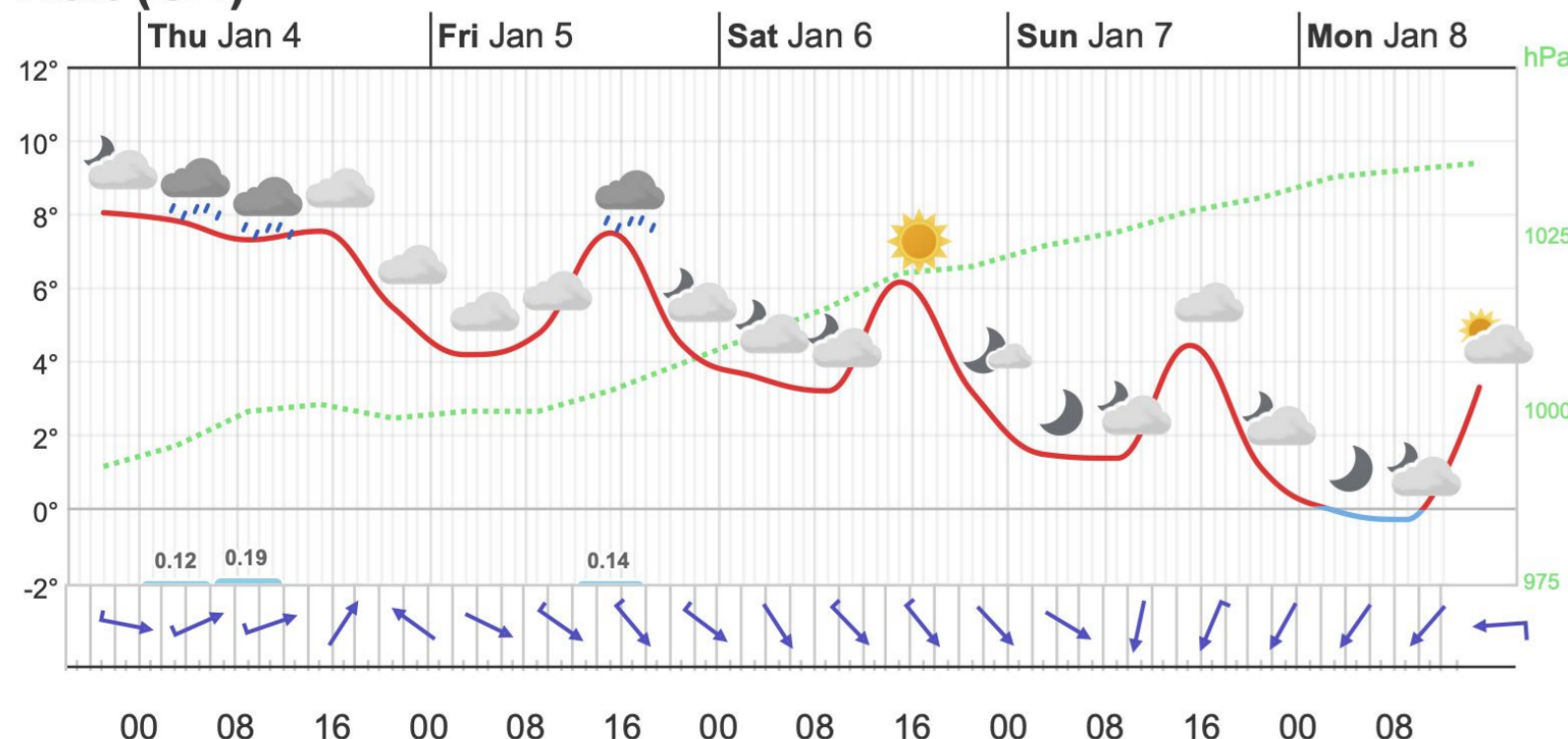


Brazil

New COVID-19 cases (smoothed), Daily new confirmed COVID-19 cases per million people, 7-d



Holt (GB)



llama2

The weather in Holt, England, will be mostly cloudy with scattered clouds for the next five days. The temperature will range from 4.2 to 7.4 degrees Celsius, with a high of 6.1 degrees Celsius on January 7th. The wind speed will be around 2.5 to 4.6 meters per second, with gusts of up to 9.32 meters per second on January 4th. There will be light rain on January 4th, with a total of 0.12 millimeters of precipitation. The humidity will be around 77% to 97%.

Future of Factgenie

More datasets, smoother workflows, faster generation!

Benefits of span error annotations

- No hard to get references
- Interpretable, explainable and extensible

Agility and easy-to-use is needed!

- ★ Factgenie streams results immediately
 - Inspect first results in seconds
 - Fast prompt/instruction annotation protocol development
- ★ Factgenie is easily customizable for one-time modification
- ★ We will be actively developing factgenie in next year
- ★ We welcome popular dataset contributions!
- ★ Share your GPU with colleagues protected by password!
 - Collaborate in prompting
 - Annotate errors yourself, using workers or LLMs
 - Inspect errors from workers, LLMs, or other colleagues

