There's a lot of insight about how students learn in education conversations. But it's painful to process and analyze. Edu-ConvoKit is an open-source library that handles all of that for you: pre-processing, annotation, and analysis.

```bash
text: pip install edu-convokit
github: https://github.com/stanfordnlp/edu-convokit
```

### Why Edu-ConvoKit?

**Motivation.** Language is central to educational interactions, ranging from classroom instruction to tutoring sessions.

**Challenges.** No centralized tool & high learning curve for performing computational analysis! From interviews:
- “What’s the best way to de-identify the data?”
- “I want an easily accessible collection of language tools that can detect insightful things.”

### Edu-ConvoKit

Edu-ConvoKit is a modular and end-to-end pipeline for:

**Pre-Processing.** The `PreProcessor` entity modifies the text, such as through deidentification. This is important to maintain the confidentiality of participants, while preserving the context of interaction.

```python
# Original data
>>> print (df)
    text 
0   My name is Alice Wang.
1   Hey Johnson, this is John.
>>> processor = TextPreProcessor(...) 
# Processed data
>>> print (df)
    text 
0   My name is [T].
1   Hey [S2], this is [S1].
```

**Annotation.** The `Annotator` entity annotates features at an utterance-level. It currently supports several measures: talk time, math density, student math reasoning, etc.

**Analysis.** The `Analyzer` entity support common analyses in education conversation research, such as the temporal analysis:

```python
>> analyzer = TemporalLayer(data) 
>> analyzer.plot() 
```

### Resources

- You can find all these resources on [https://github.com/stanfordnlp/edu-convokit](https://github.com/stanfordnlp/edu-convokit)
- Need a step-by-step walkthrough of the features? We have Colab notebooks for each entity.
- Need examples of Edu-ConvoKit applied to real datasets? We have Colab notebooks where we applied Edu-ConvoKit to NCTE (classroom), TalkMoves (classroom) and Amber (tutoring) datasets.
- Need references for Edu-ConvoKit’s features and applications? We have a paper database on our GitHub repository.
- Need library documentation? Yes, we gotchu too! [https://edu-convokit.readthedocs.io/](https://edu-convokit.readthedocs.io/)
- Can find a feature that you need? Contribute to Edu-ConvoKit or reach out!

---

**Email:** rewang@cs.stanford.edu  
**Twitter:** rose_e_wang