

## Overview

- Software-based energy measurements are inaccurate (Cao et al, 2020)
- IrEne-viz presents an interactive demonstration of energy consumption of different models and their components

## IrEne-viz can help

- Identify specific bottlenecks in a model in order to increase energy efficiency
- Tailor models for a specific use case e.g., battery-powered mobile devices

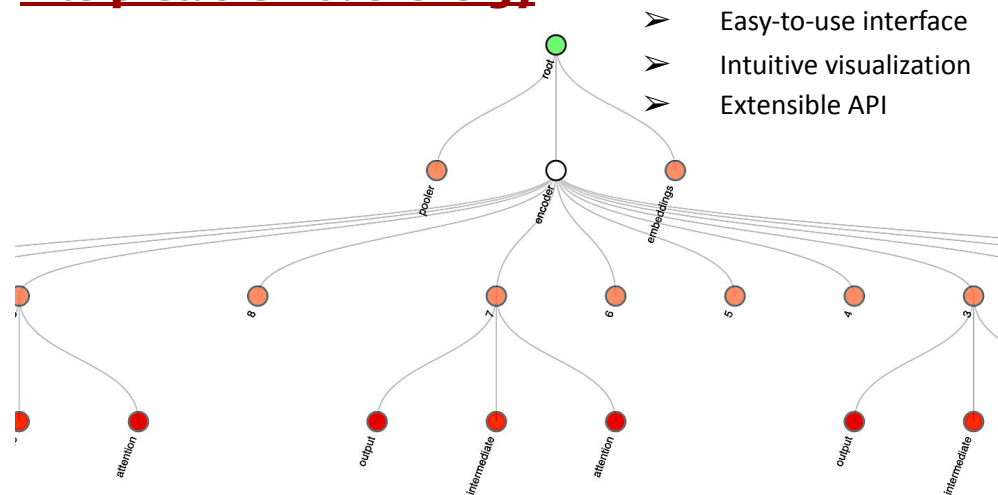
## Energy Information

Name: root.encoder.6.attention
Predicted energy: <b>189.326 mJ</b>
Instance Type: robertaattention
Memory (bytes): 189.750
CPU: 5.662
Flops: 2472.739
GPU: 94.000
GPU memory: 19.846
GPU energy mean: 125.260 mJ
Level: 3
Level type: module
Level % Energy : 127.72
Model % Energy : 3.64

## Identify Bottlenecks

Node Name	Pred. Energy (mJ)
<a href="#">transformer</a>	2015.527
<a href="#">transformer.5</a>	307.248
<a href="#">transformer.0</a>	307.247
<a href="#">transformer.1</a>	307.247
<a href="#">transformer.2</a>	307.247
<a href="#">transformer.3</a>	307.247

## Interpretable model energy



## Check out the Demo!

Or, come find us  
[stonybrooknlp.github.io/irene/demo](https://stonybrooknlp.github.io/irene/demo)

## References

Cao et al, 2020. *Towards Accurate and Reliable Energy Measurement of NLP Models*, SustainNLP 2020