Cogito Ergo *Summ*: Abstractive Summarization of Biomedical Papers via Semantic Parsing Graphs and Consistency Rewards

ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA

Giacomo Frisoni, Paolo Italiani, Stefano Salvatori, Gianluca Moro

Department of Computer Science and Engineering, University of Bologna, Italy

Paper

Motivations

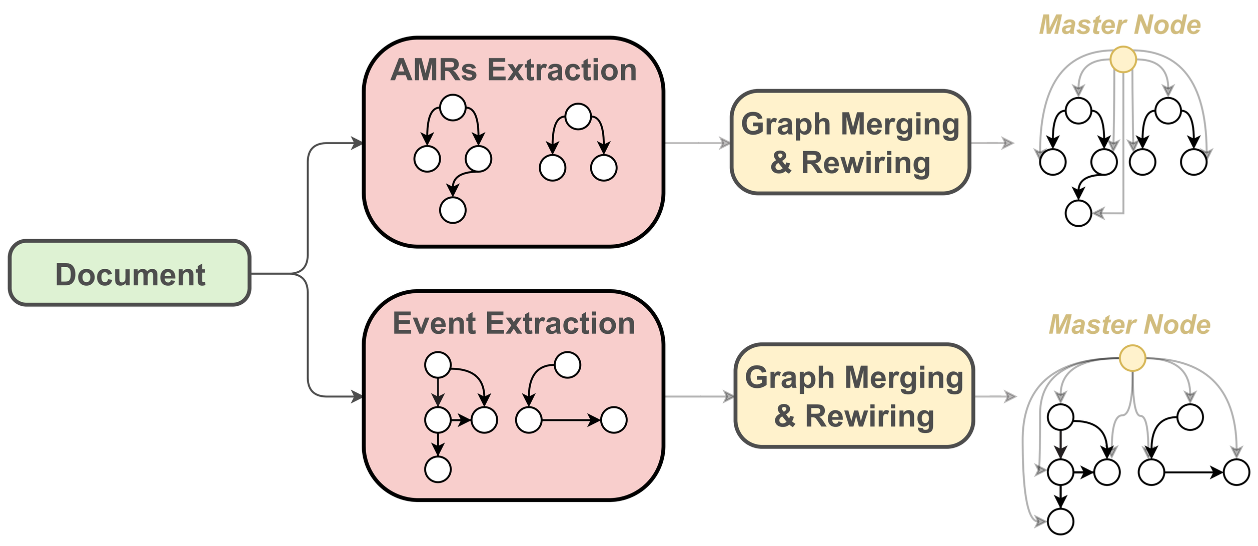
- Biomedical document summarization challenges
  - Medical jargon truly hard to interpret
  - Clauses' interdependence, complex interactions, precise domain information, narrow interpretation margin, no factual mistakes
- Problems and weaknesses of existing solutions
  - Highly prone to hallucinating content or falling back on extraction
  - Superficial text organization rather than underlying semantics
  - Not ensuring document-summary consistency

Contribution

- **CogitoErgoSumm**, the first semantics-aware transformer-based model for single document abstractive summarization in the biomedical domain
  - Combining pre-trained language models and semantic parsing graphs providing formal meaning representations
  - Two different semantic parsing techniques with complementary strengths: **Event Extraction (EE)** and **Abstract Meaning Representation (AMR)**
- **Reinforcement Learning (RL)** to ensure factuality and consistency
  - Reward function based on the average SMATCH score between the original document and the generated summary

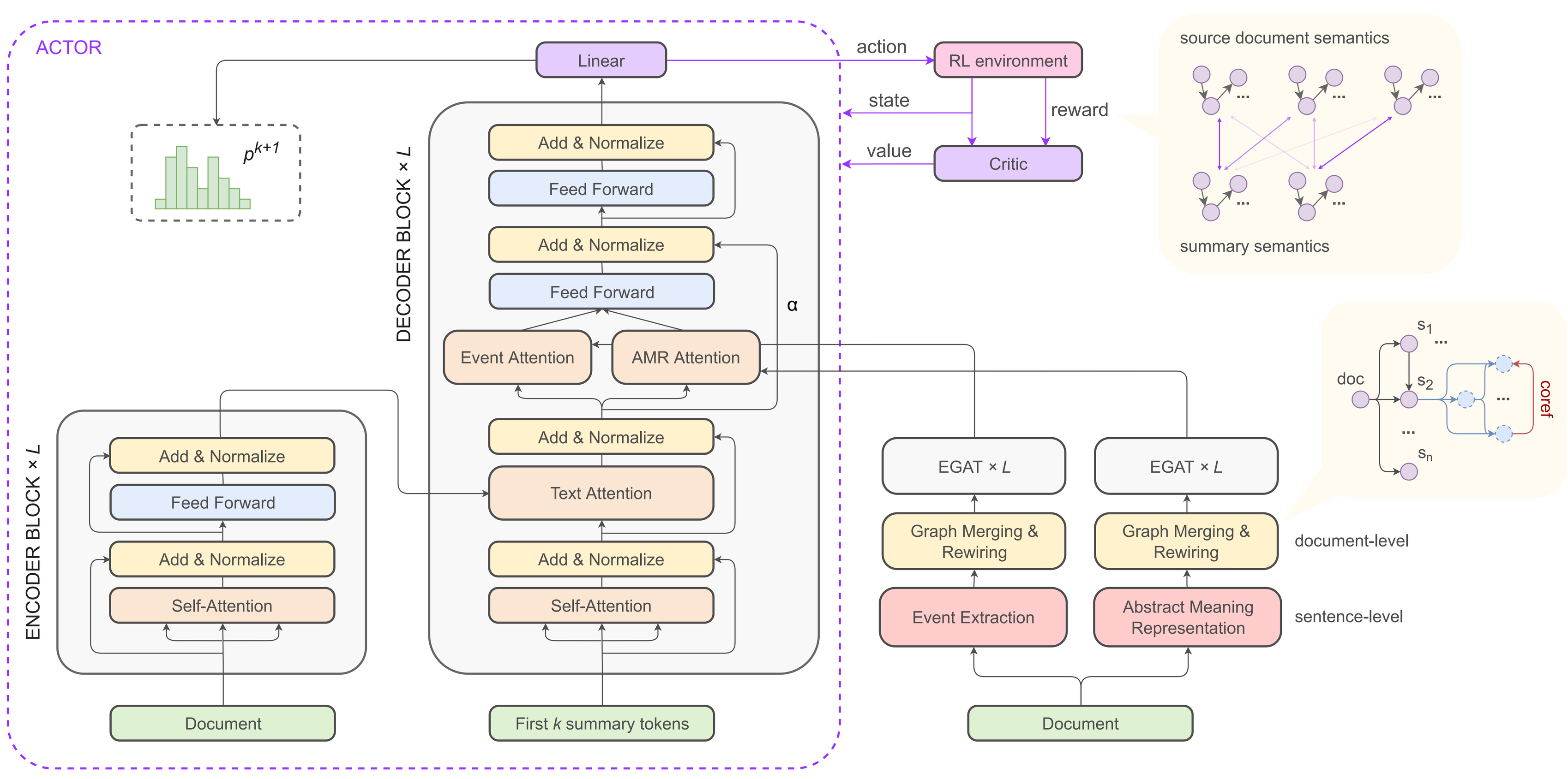
AMR and Event Graphs

- **Abstract Meaning Representation Graphs**
  - Capture the general meaning of any sentence as high-level semantic relations (abstraction from words to concepts)
  - We use SPRING to automatically extract AMRs from sentences
- **Event Graphs**
  - Capture biomedical-specific interactions with n-ary and potentially nested interactions between participants
  - We use DEEPEVENTMINE to automatically extract events



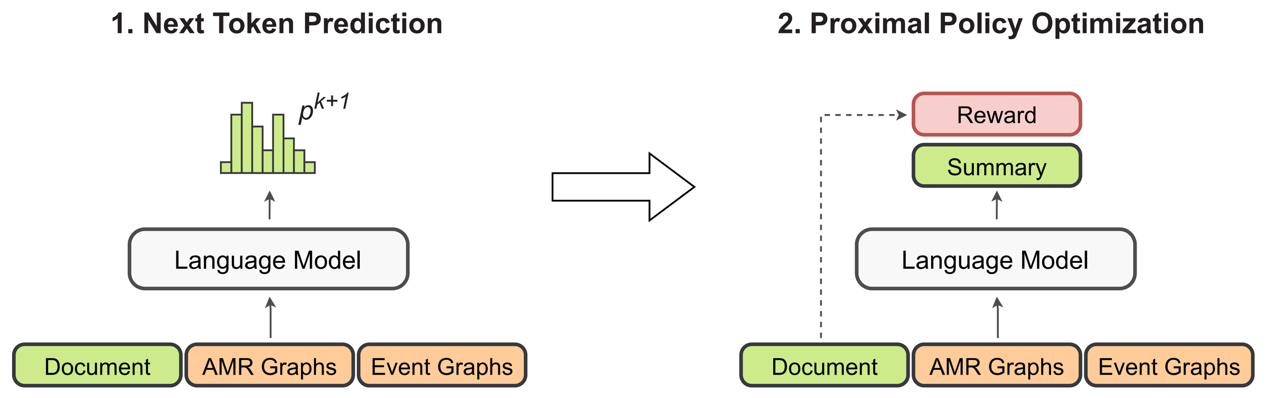
Architecture

- We extend a pre-trained BART-base architecture with the nimble ability to **attend semantic parsing graphs** during decoding and preserve the most relevant information via **RL**



Method

- Two training phases
  - 1^{th} Phase → Next Token Prediction, but with semantic graphs
  - 2^{th} Phase → RL with *Proximal Policy Optimization*



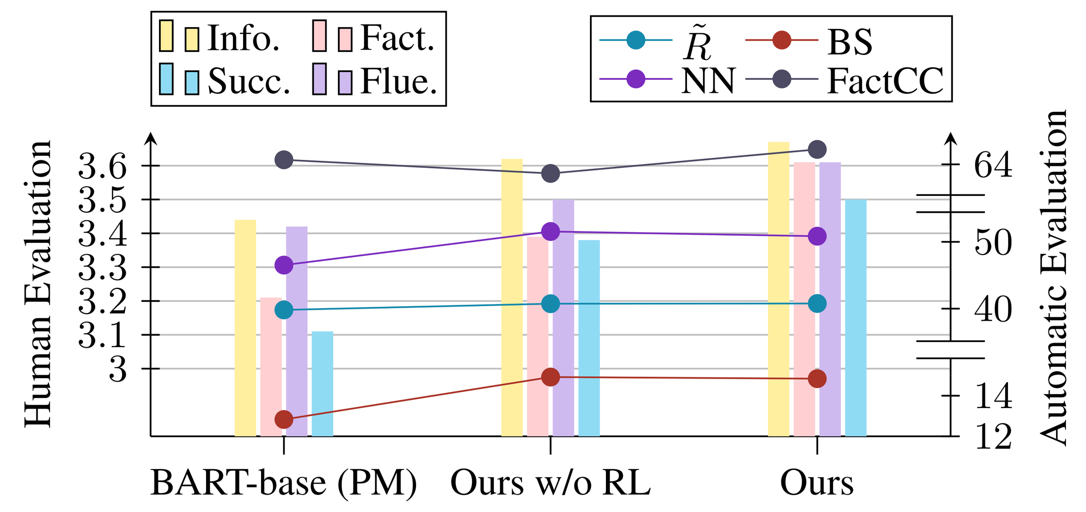
- We add **two extra cross-attentions layers** to the BART decoder: *Event Attention* and *AMR Attention*
  - Node representations obtained using edge-aware GAT layers
- We use **RL to preserve as much pivotal information as possible** from the original document
  - Maximize the overlap between the AMRs of the input document and the AMRs of the generated summary with SMATCH
  - Try not to deviate too much from the pr-etrained model by keeping a low KL-divergence

$$\psi(doc, summ) = AvgSmatch(doc, summ) - \beta \log \frac{\pi_\theta(a_t|s_t)}{\pi_{base}(a_t|s_t)}$$

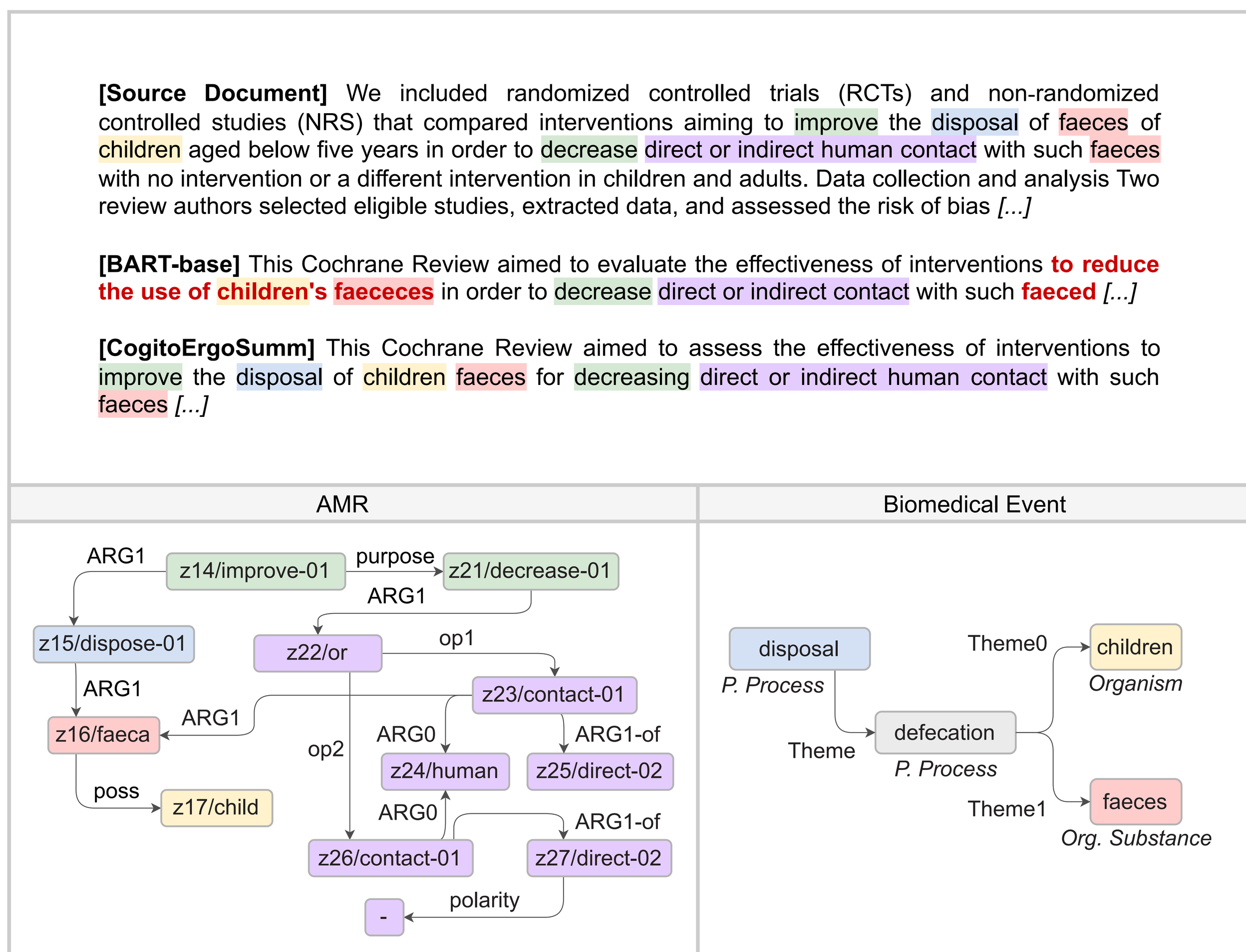
Results

Model	#params	R-1	R-2	R-L	Flesch-Kincaid	Coleman-Liau
ORACLE†	—	53.56	25.54	49.56	14.85	16.13
BERT-base†	110M	26.60	11.11	24.59	13.44	14.40
POINTER GENERATOR†	22M	38.33	14.11	35.81	16.36	15.90
BART-base (PubMed)	139M	51.20	19.77	48.47	13.69	13.45
BART-large (PubMed)†	406M	**52.66**	**21.73**	**49.97**	13.30	14.28
EASUMM‡	8M	46.30	18.73	43.78	12.42	13.06
COGITOERGOSUMM	181M	52.23	20.63	49.44	14.10	13.67
- w/o RL	180M	52.30	20.47	49.46	14.06	13.64
- w/o event and RL	155M	52.13	20.42	49.30	14.02	13.69
- w/o AMR and RL	157M	52.02	20.54	49.25	13.97	13.66

- ROUGE scores higher than most of the previous methods
- Still competitive with BART-large despite having 2x fewer parameters



- Better results on every quality dimension in the human evaluation (+12.46% factualness, +6.69% informativenes)
- The plot also underlines the poor correlation between ROUGE and the desired output properties



- Qualitative example of induced semantic parsing graphs and their assistance to high-quality summarization

Contacts: {giacomo.frisoni, paolo.italiani, s.salvatori, gianluca.moro}@unibo.it