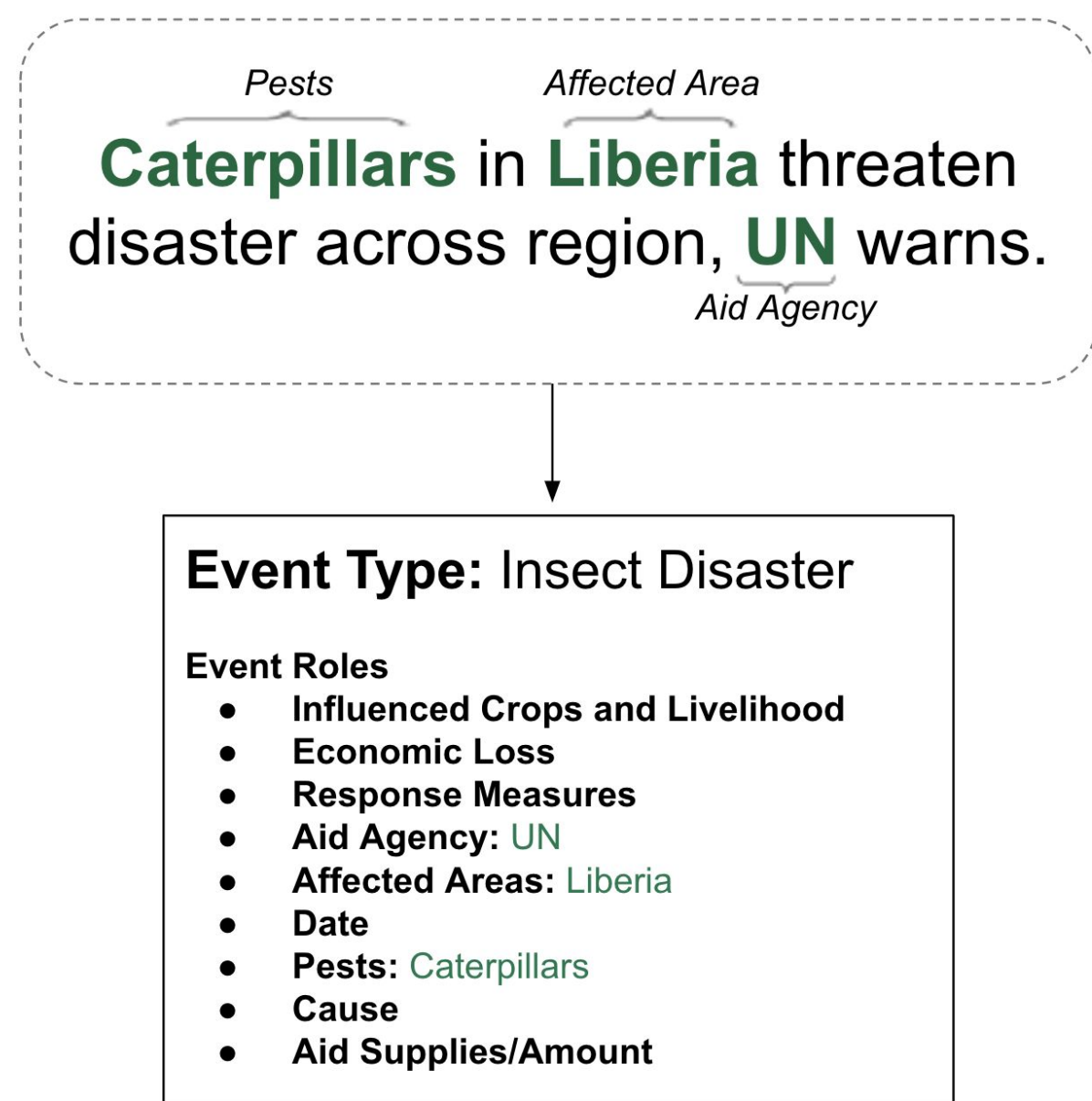# Document-Level Event-Argument Data Augmentation for Challenging Role Types
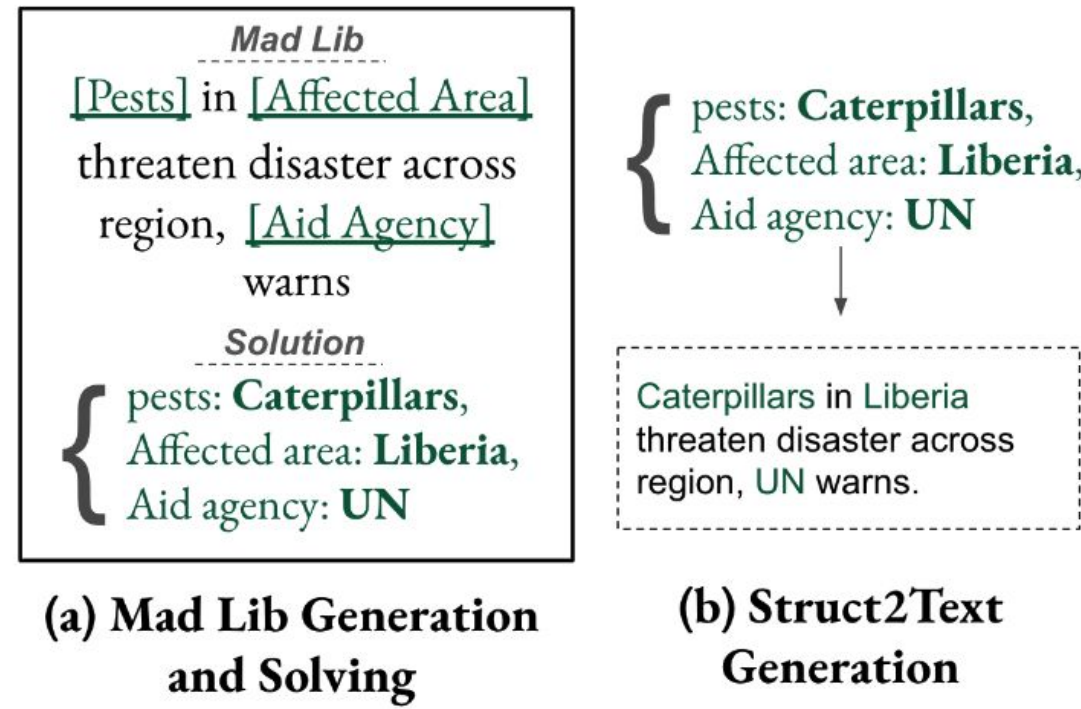
Joseph Gatto, Omar Sharif, Parker Seegmiller, Sarah Preum
Dartmouth College, Department of Computer Science

Pests   *Affected Area*
Caterpillars in **Liberia** threaten disaster across region, **UN** warns.
Aid Agency

Event Type: Insect Disaster

Event Roles
- **Influenced Crops and Livelihood**
- **Economic Loss**
- **Response Measures**
- **Aid Agency:** UN
- **Affected Areas:** Liberia
- **Date**
- **Pests:** Caterpillars
- **Cause**
- **Aid Supplies/Amount**

Example of the Event Argument Extraction (EAE) task

PROBLEM

- Event argument extraction (EAE) from long documents is *very challenging*.
- Document-Level EAE (DocEAE) data collection for new domains is *extremely expensive*.
- Goal: We want to synthesize novel documents for DocEAE training

Mad Lib
[Pests] in [Affected Area] threaten disaster across region, [Aid Agency] warns.

Solution
pests: **Caterpillars**, Affected area: **Liberia**, Aid agency: **UN**

{ pests: **Caterpillars**, Affected area: **Liberia**, Aid agency: **UN** }

Caterpillars in Liberia threaten disaster across region, UN warns.

(a) Mad Lib Generation and Solving   **(b) Struct2Text Generation**

Key: LLMs struggle to synthesize long documents (10+ sentences) for span annotation tasks like DocEAE!

OUR SOLUTION

Novel Inference Framework: We introduce two methods for DocEAE data generation that meet the following criteria:
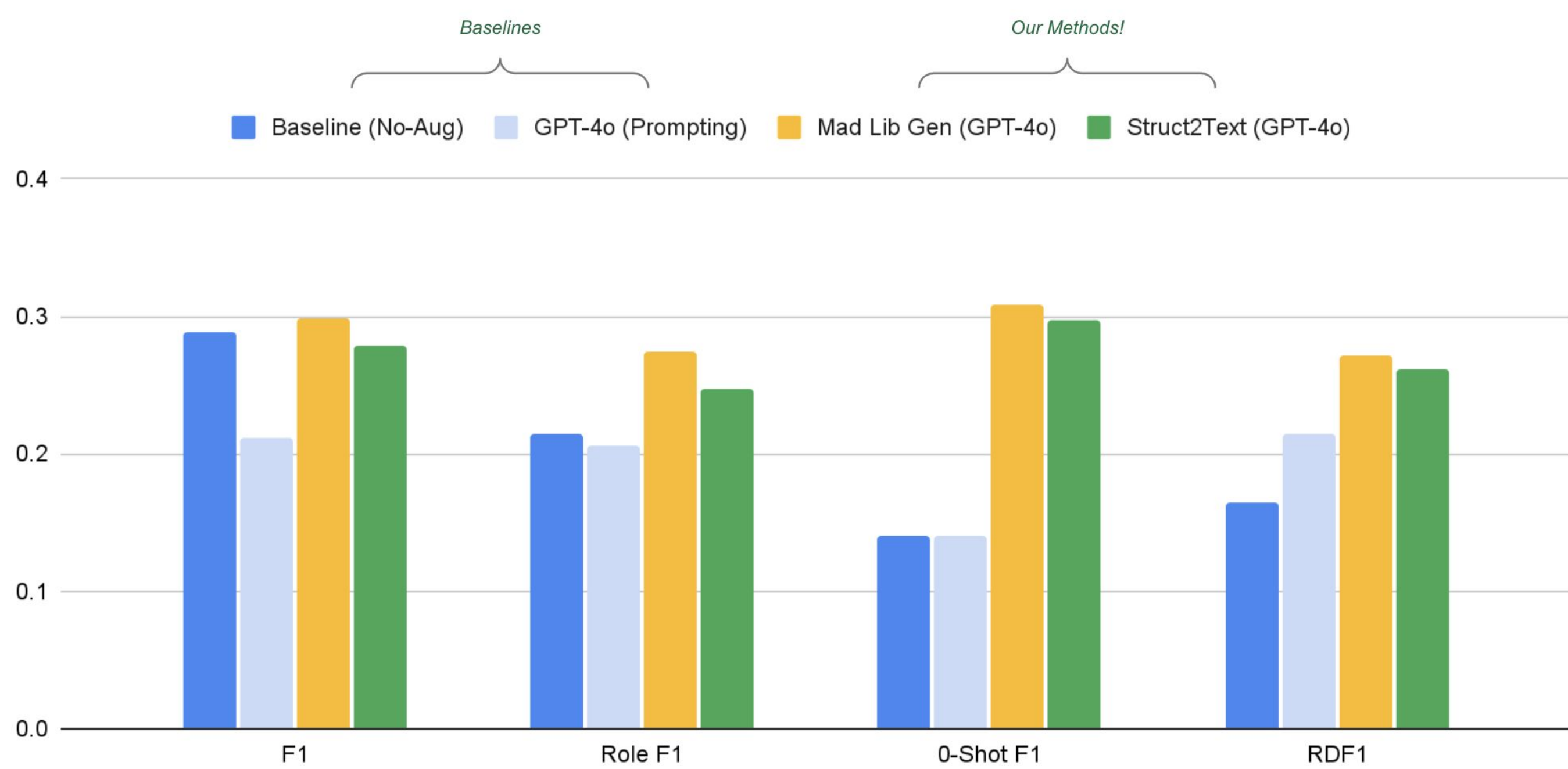
☐ **Cross-Domain Generation:** Our methods use zero in-domain training samples to create new training data, relying on out-of-domain data + LLM parametric knowledge. This enables generation of long-tail role types.

☐ **LLM Span Generation Alignment:** Each of our methods ensures the desired event argument has accurate span annotation in the synthesized document.

Methods: We can generate novel EAE samples in *low-resource domains* using *zero in-domain data*.
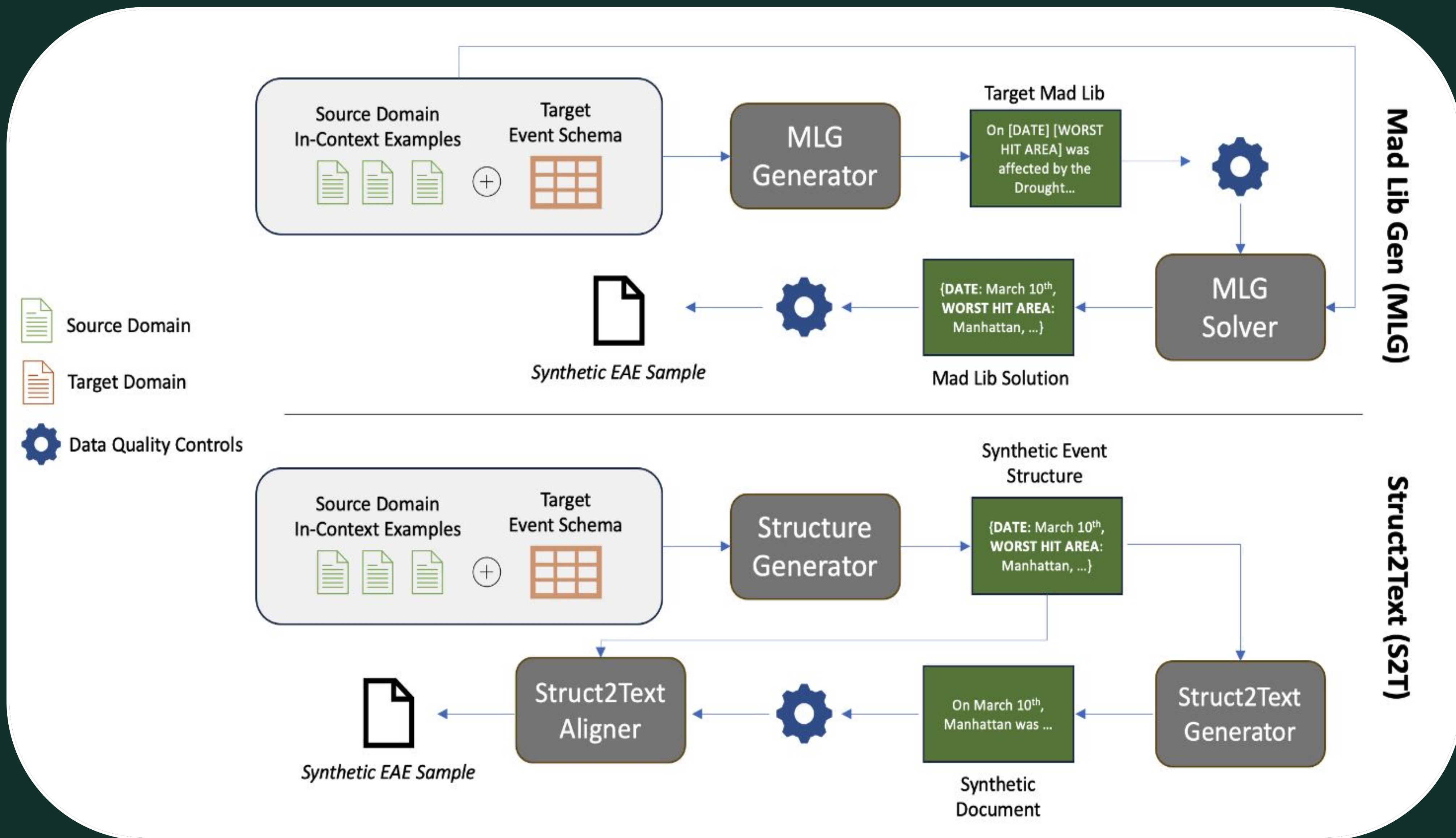
Experiment #1: Can we improve few-shot, cross-domain performance on document-level EAE?

Metric	Description
F1	Standard F1 On All Data
Role F1	Compute F1 For *Each Role Individually*, Then Take Average
0-Shot F1	Role F1 for 0-Shot Roles (n=9)
Role-Depth F1 (RDF1)	Role F1 for **Semantically Outlying Roles** (n=16)

Result #1: On the DocEE few-shot cross-domain data split, both methods provide significant improvement on 0-shot and semantically outlying role types!

Experiment #2: Does MLG improve performance on standard EAE tasks?

	F1	Role F1
DiscourseEE (10%)	0.13	0.059
DiscourseEE (10%) + Aug	0.334[†]	0.342[†]
DiscourseEE (50%)	0.163	0.092
DiscourseEE (50%) + Aug	0.361[†]	0.357[†]
DiscourseEE (Full)	0.18	0.106
DiscourseEE (Full) + Aug	0.403[†]	0.396[†]
RAMS (10%)	0.134	0.128
RAMS (10%) + Aug	0.214[†]	0.186[†]
RAMS (50%)	0.323	0.298
RAMS (50%) + Aug	0.343[†]	0.33[†]
RAMS (Full)	0.388	0.38
RAMS (Full) + Aug	0.393	0.375
PHEE (10%)	0.42	0.303
PHEE (10%) + Aug	0.544[†]	0.53[†]
PHEE (50%)	0.596	0.581
PHEE (50%) + Aug	0.599	0.594
PHEE (Full)	0.621	0.618
PHEE (Full) + Aug	0.618	0.608

Result #2: Our data augmentation strategy improves low-resource performance on DiscourseEE, RAMS, and PHEE.

DARTMOUTH
Guarini School of Graduate and Advanced Studies