Weakly Supervised Data Augmentation through Prompting for Dialogue Understanding

Maximillian Chen, Alexandros Papangelis, Chenyang Tao, Andy Rosenbaum, Seokhwan Kim, Yang Liu, Zhou Yu, Dilek Hakkani-Tur

COLUMBIA | ENGINEERING
The Fu Foundation School of Engineering and Applied Science

amazon | science

Background & Challenges

High quality conversational data is scarce.

- **Data augmentation** techniques can be used to expand limited training data
- **Prompting large pre-trained language models** appears to be a viable method of generating coherent synthetic dialogue turns
- But: uncontrolled generation means no guarantee on utterance "correctness"

(neutral) Alice: You're going to set up your own law office, aren't you?
(neutral) Bob: Yes. After so many years of hard work, I'd rather I had an office of my own.
(happy) Alice: If you need help, don't hesitate to ask me.
(happy) Bob: I'll be very glad if you would help.
(happy) Alice: I'd like to wish you every success in your new venture.
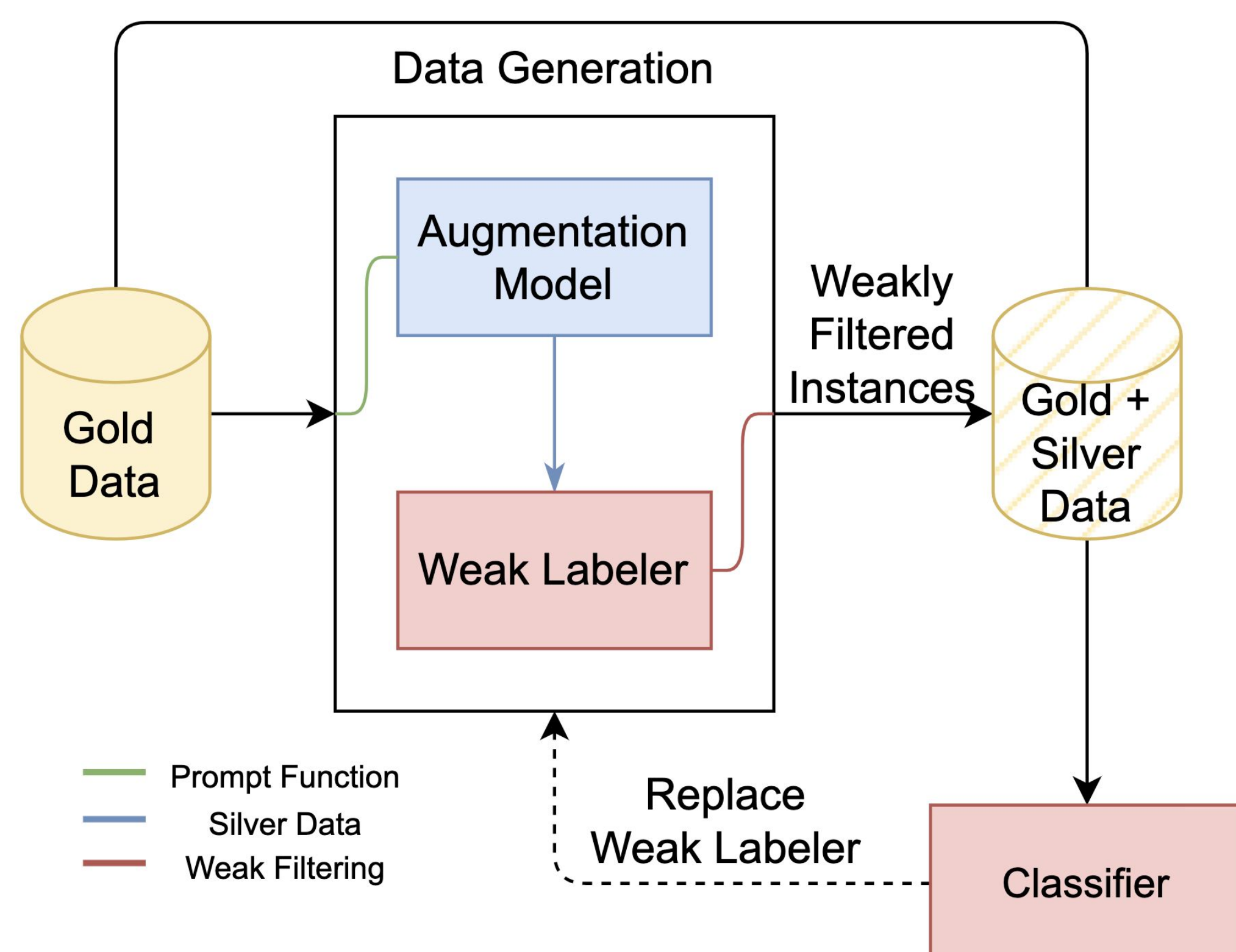(happy) Bob: Thank you. I wish I would.
(happy) Alice:
Generated Responses:
1. Good luck to you. Let's do lunch soon, Bob.
2. It's such a rare pleasure to meet such an ideal partner in your work.
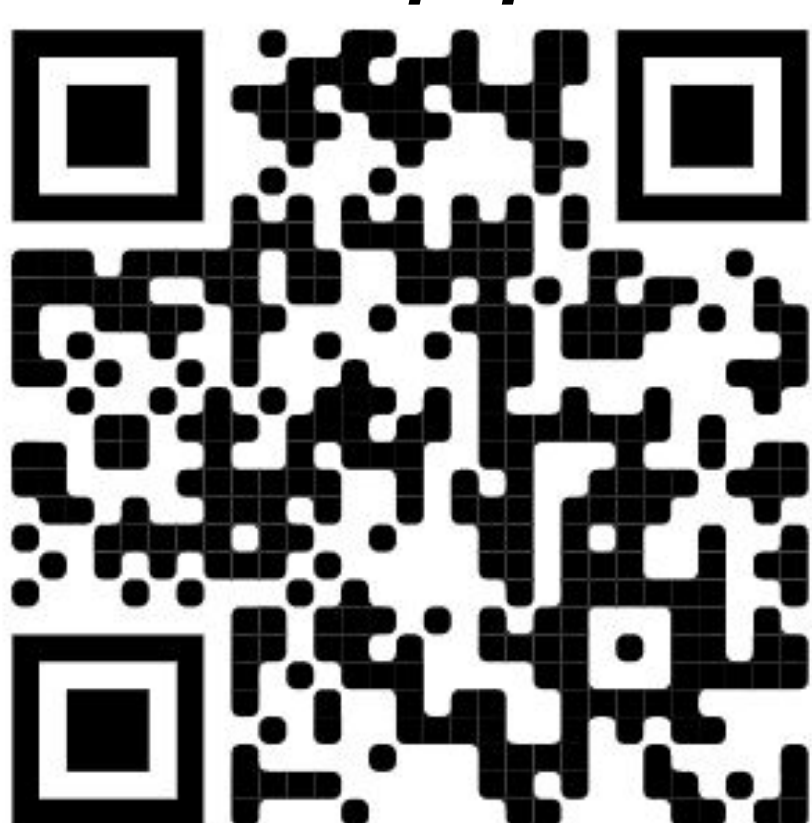3. You know, you seem quite different.

Refining Data with Weak Supervision

WeakDAP workflow:
- Prefix prompt to generate synthetic turns
- Weakly classify each prospective synthetic instance's label, rejecting silver data if mismatching desired label with high confidence
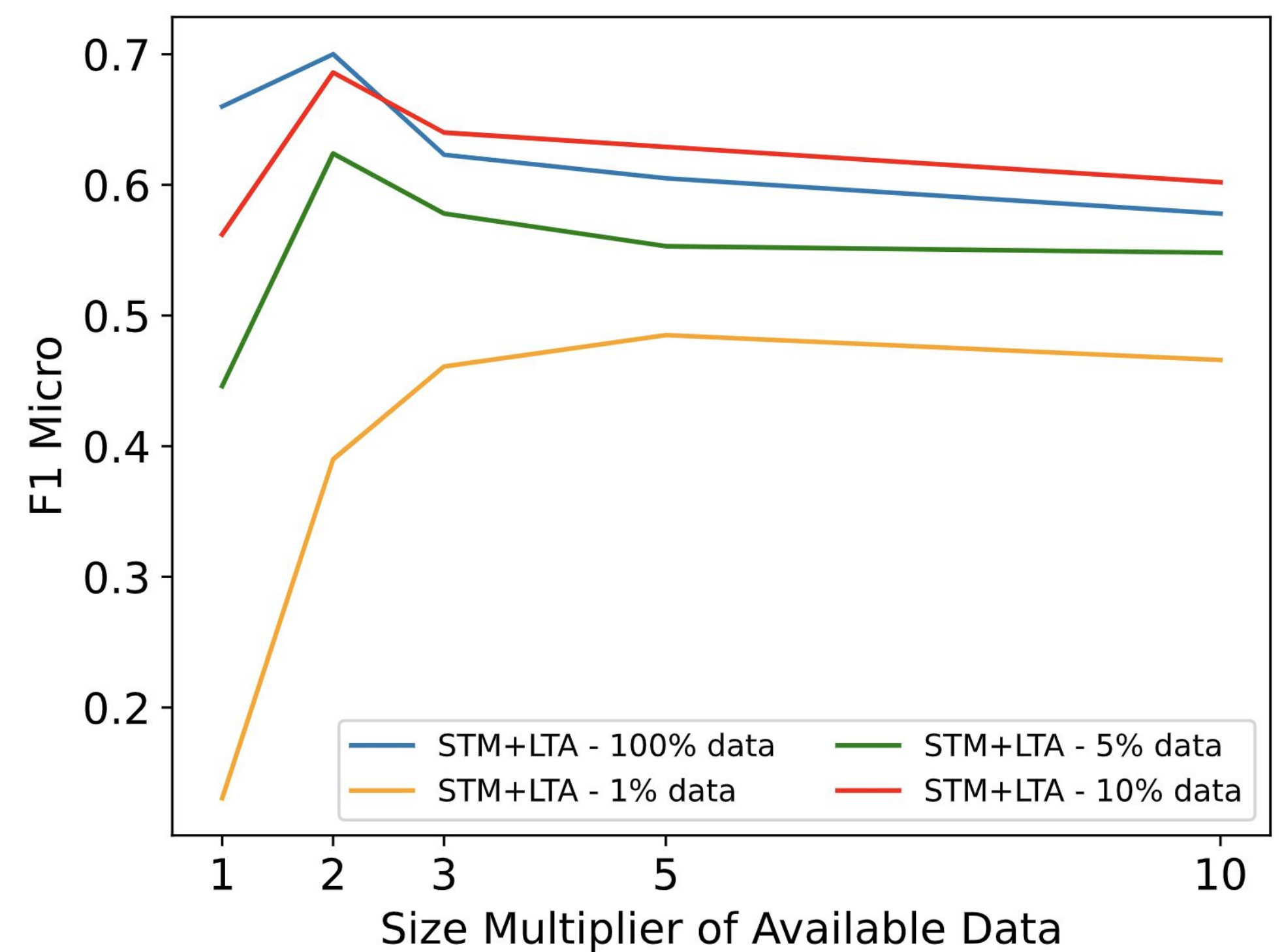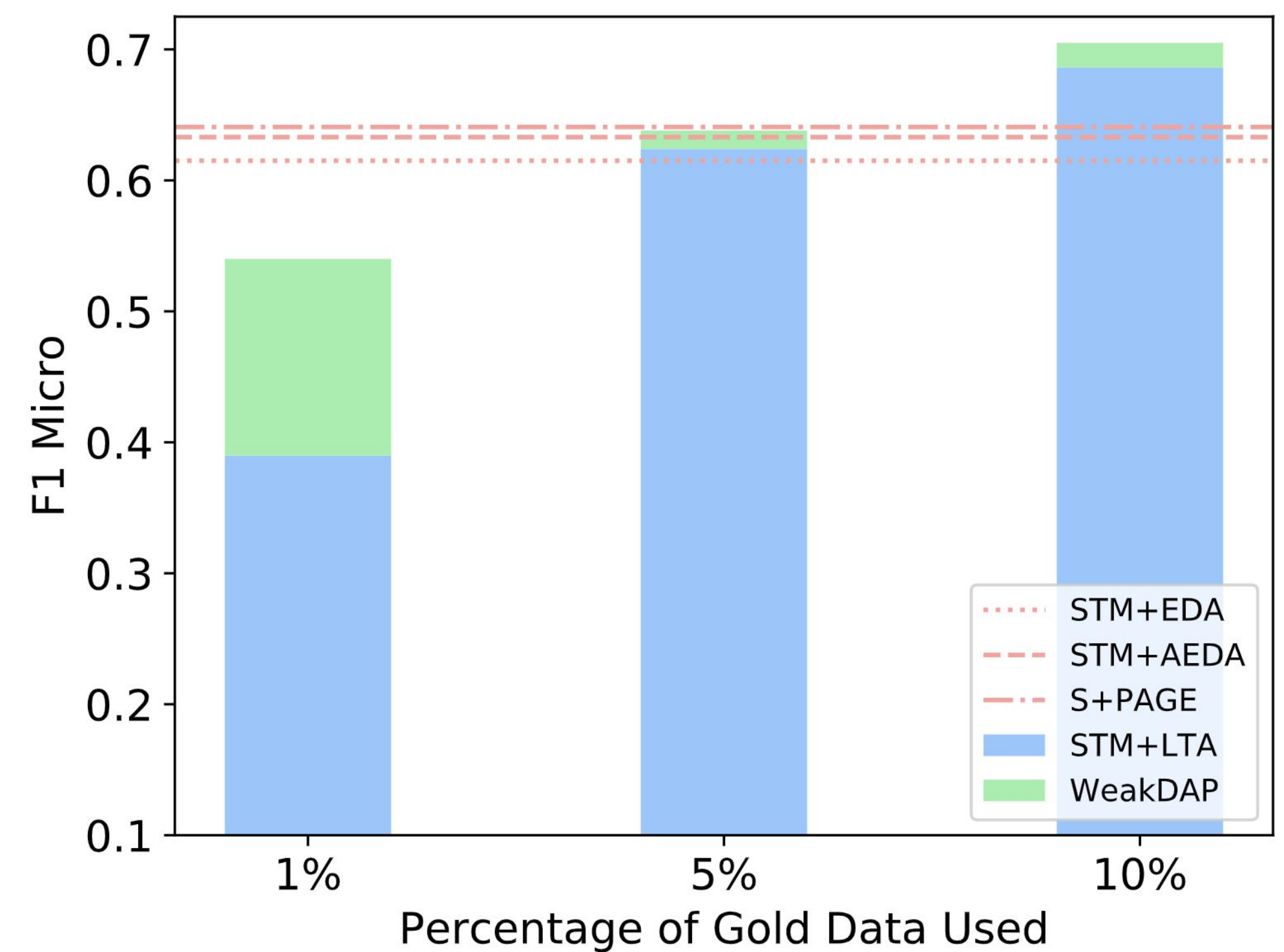- Repeat until convergence

Link to paper!

WeakDAP can be adapted to any large pre-trained large language model, and any classifier for a given dialogue understanding task

Results: Emotion Classification

- Varied seed and synthetic data sizes in DailyDialog emotion classification task
- Prompting yields large improvements at 2x

- Applying WeakDAP improves all cases
- Surpasses full-shot SotA with 10% gold data

Results: Cross-lingual Intent Detection

FBTOD ES	Acc_{LR}	$F1_{LR}$	Acc_{HR}	$F1_{HR}$
1% No Aug.	0.572	0.164	0.572	0.164
1% Prompt	0.737	0.316	0.776	0.359
+ WeakDAP	**0.834**	**0.495**	**0.831**	**0.528**
5% No Aug.	0.845	0.417	0.845	0.417
5% Prompt	0.953	0.641	0.954	0.682
+ WeakDAP	**0.957**	**0.715**	**0.962**	**0.732**
10% No Aug.	0.942	0.588	0.942	0.588
10% Prompt	0.973	0.772	0.973	0.791
+ WeakDAP	**0.979**	**0.905**	**0.976**	**0.846**
100% No Aug.	0.988	0.889	0.988	0.889

- Few-shot Spanish Intent Detection - FBTOD
- Augmentation using mixed Spanish and English in-context examples
- HR/LR = High/Low-Resource English in-context example pool