Рассуждение с помощью больших языковых моделей в задаче анализа тональности на русском языке

Nicolay Rusnachenko

nicolay-r.github.io

Outline

1. 📊 How well **Large Language Models can actually reason** *as-it-is* depending on scales in Sentiment Analysis
2. 🔗 How to apply **Chain-of-Thought** in Sentiment Analysis
3. 🔥 **How to build** the most-recent advanced Sentiment Analysis system

Sentiment Analysis

Is an extraction of author opinion towards objects mentioned in text:

Opinions

Author
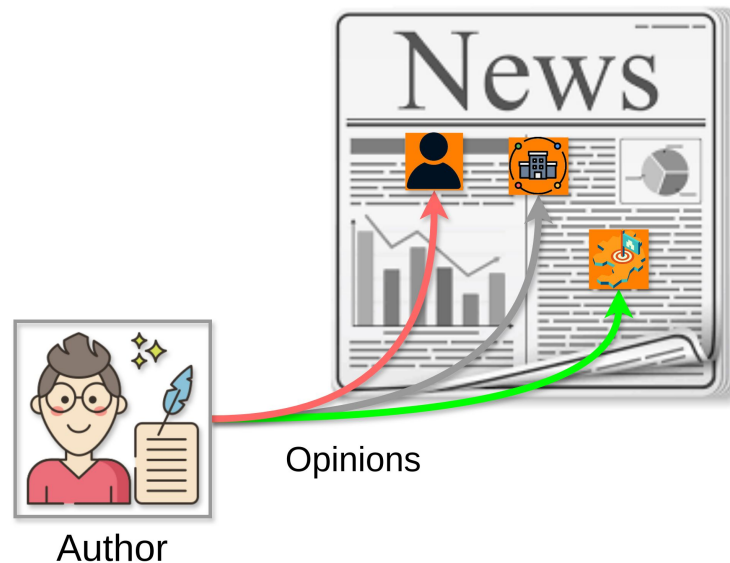
Keynotes about the task:

💡 Potentially huge amount of mentioned **named entities.**

💡 Documents may vary in size ➡ complexity of the task

Text Classification

P. Turney 2004 — Given text, the task is to define to which sentiment class it is related to: positive, negative, neutral

Президент Цай Инвэнь заявила, что Китай использует **финансовое давление** и пытается повлиять на результаты предстоящих президентских выборов на Тайване.

Limitations:

💡 Presume that we deal with the single entity in text

💡 Texts are expected to be short
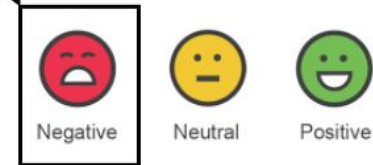
Targeted Sentiment Analysis

Президент **Цай Инвэнь** заявила, что **Китай** использует финансовое давление и пытается повлиять на результаты предстоящих президентских выборов на **Тайване**.

negative

Author

Towards Large Language Models 🤖

Towards Large Language Models 🤖 (Architecture)

Conventional Methods

Neural Networks

- CNN
- LSTM

Self-Attention

Transformers (LM)

Encoders

BERT

Decoders

GPT, T5
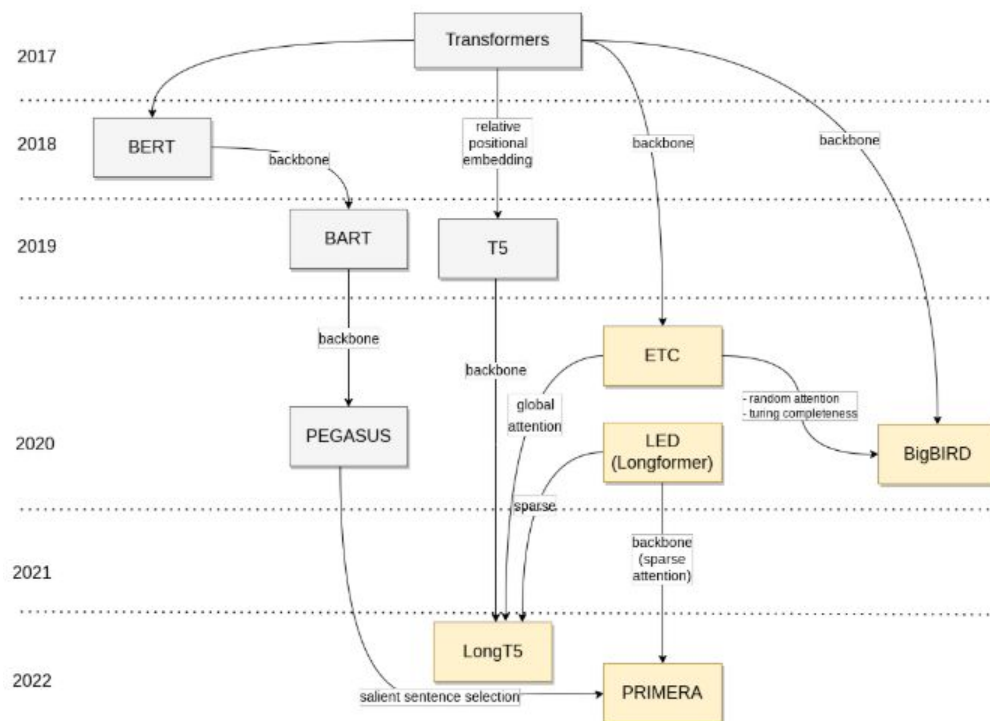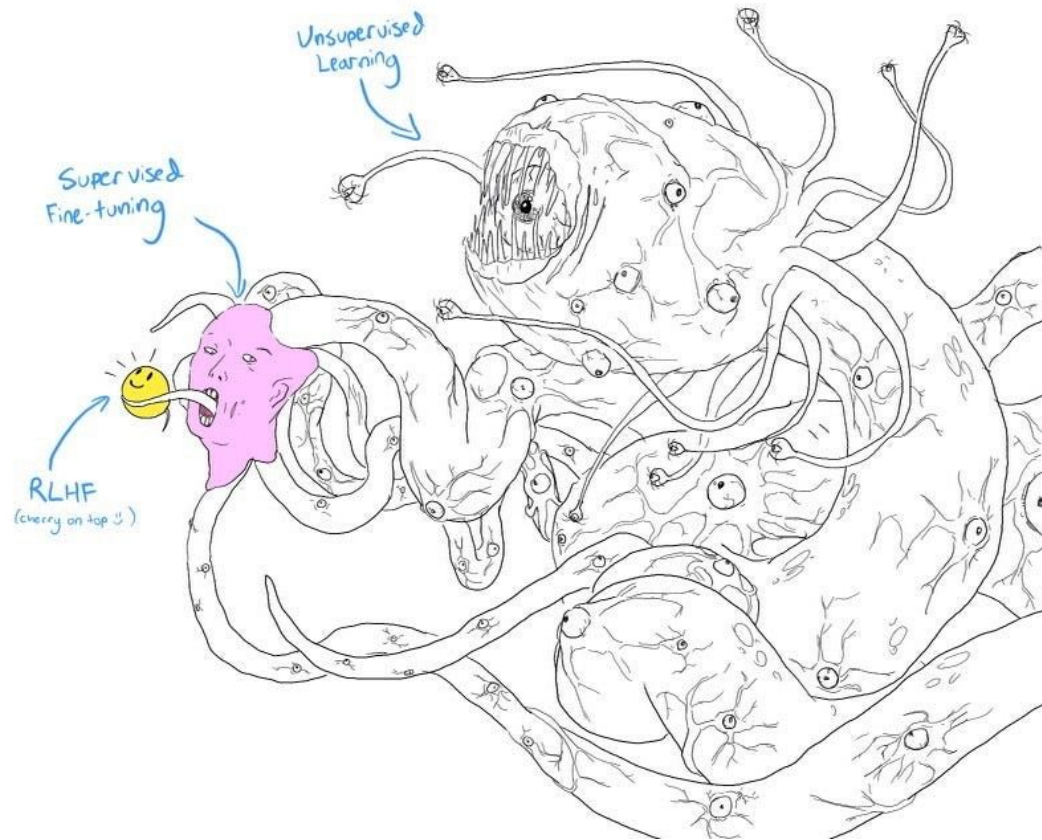
Towards Large Language Models 🤖 (Training)

List the main training techniques:

1. Unsupervised Pre-training
   a.   + Distant Supervision (optional)
2. Supervised Tuning
   a. **Instructions**
3. Reinforcement Learning With Human Feedback (*RLHF* Framework)

LLM: *LM + Instruction tuning + (RLHF)*

Unsupervised Learning

Supervised Fine-tuning

RLHF
(cherry on top 😋)

Experiments 🧪

Experiments 🧪 1/2 (Dataset Setup)

RuSentNE-2023 — sentences from Wikinews articles (CC BY 2.5), annotated mentioned named entities. Content from 400 articles.

Entity types: ORG, PER, LOC, COUNTRY, PROFESSION, NATIONALITY

Three classes: Positive, Negative, Neutral

Splits Statistics:

1. Train (10K)
2. **Validation (2.8K)**
3. Test (1.9K)

Evaluation Metric: F1(PN), F1(PN0)

Two versions:

1. Original Texts (Russian)

sentence	Entity Value	Entity Type	label
25 июля 2016 года одна из двух крупнейших телекоммуникационных компаний США Verizon Communications официально объявила о покупке основного бизнеса одного из крупнейших мировых порталов Yahoo!	Yahoo!	ORGANIZATION	1
Формально все решения принимались коллективно, однако эксперты именно на Ли Пэн возлагают ответственность за кровавые репрессии.	Ли Пэн	PERSON	-1
«Нотр-Дам — исторический символ Франции, бесценное сокровище европейской и мировой культуры, одна из важнейших христианских святынь.	Франции	COUNTRY	0
В основном действующими лицами рассказов Элис выступают молодые канадские девушки, со всеми трудностями и проблемами при взрослении, со страданиями и одиночеством стареющих женщин.	Элис	PERSON	0

2. Translated Texts (English) (via Googletrans)

| sentence	Entity Type	Entity Value	label
On July 25, 2016, one of the two largest US telecommunications companies, Verizon Communications, officially announced the purchase of the core business of one of the world's largest portals Yahoo!	Yahoo!	ORGANIZATION	1
Formally, all decisions were made collectively, but experts hold Li Peng responsible for the bloody repressions.	Li Peng	PERSON	-1
"Notre Dame is a historical symbol of France, a priceless treasure of European and world culture, one of the most important Christian shrines.	France	COUNTRY	0
Basically, the characters in Alice's stories are young Canadian girls, with all the difficulties and problems of growing up, with the suffering and loneliness of aging women.	Alice's	PERSON	0

Experiments 🧪 2/2 (Zero-Shot-Learning mode)

Ru: *Каково отношение автора или другого субъекта в предложении {**Sentence**} к {target}? Выбери из трех вариантов: позитивная, негативная, нейтральная*

En: *What is the attitude of the author or another subject in the sentence {**Sentence**} to the target {target}? Choose from: positive, negative, neutral.*

List of the models

Open:

- 3-4B   *tiny-sized* Microsoft-Phi
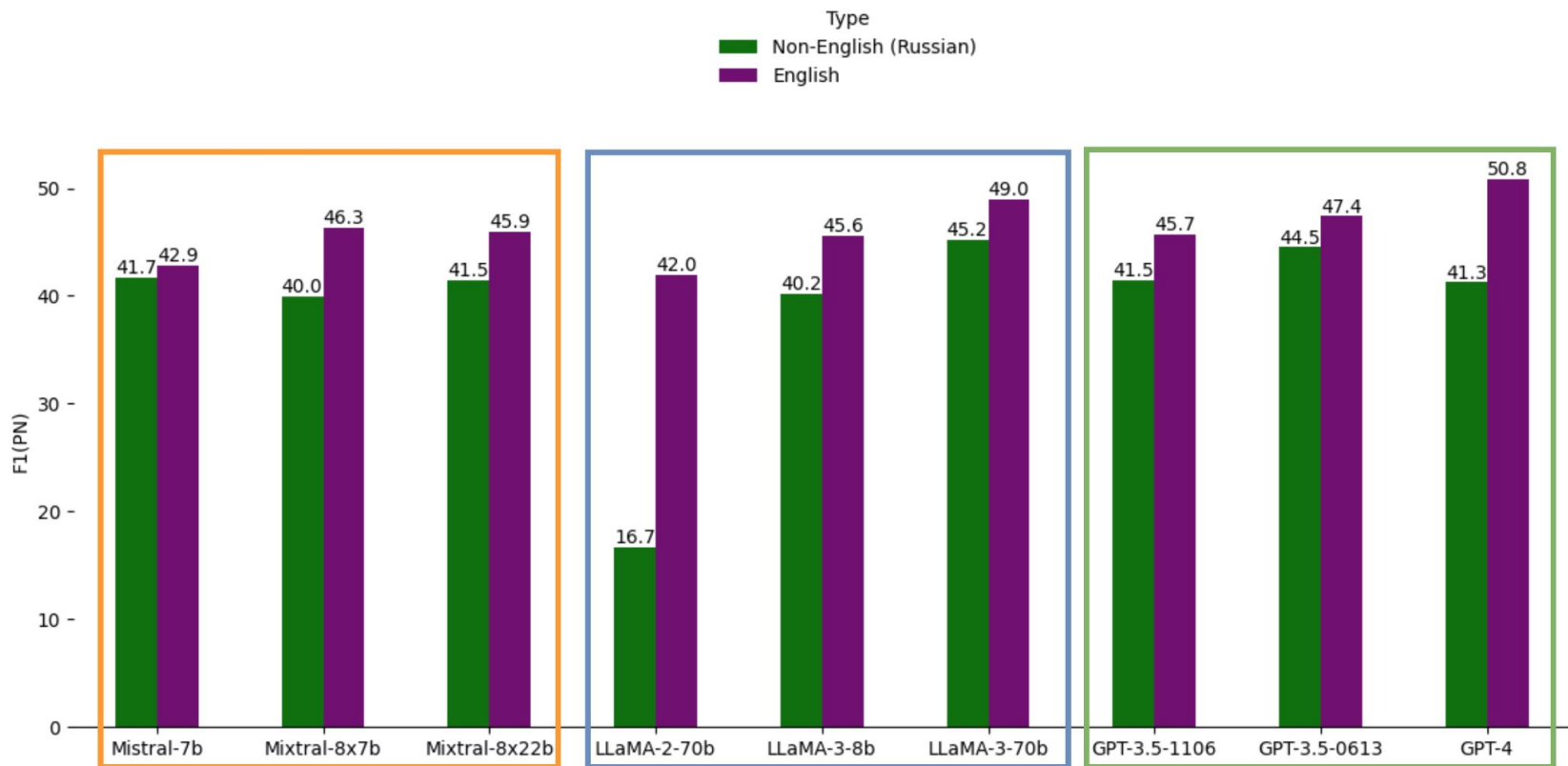- 7B   *small-sized*
- 56-70B   *medium*
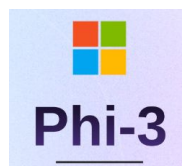
Proprietary:

- 175B   ChatGPT-3.5
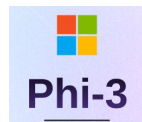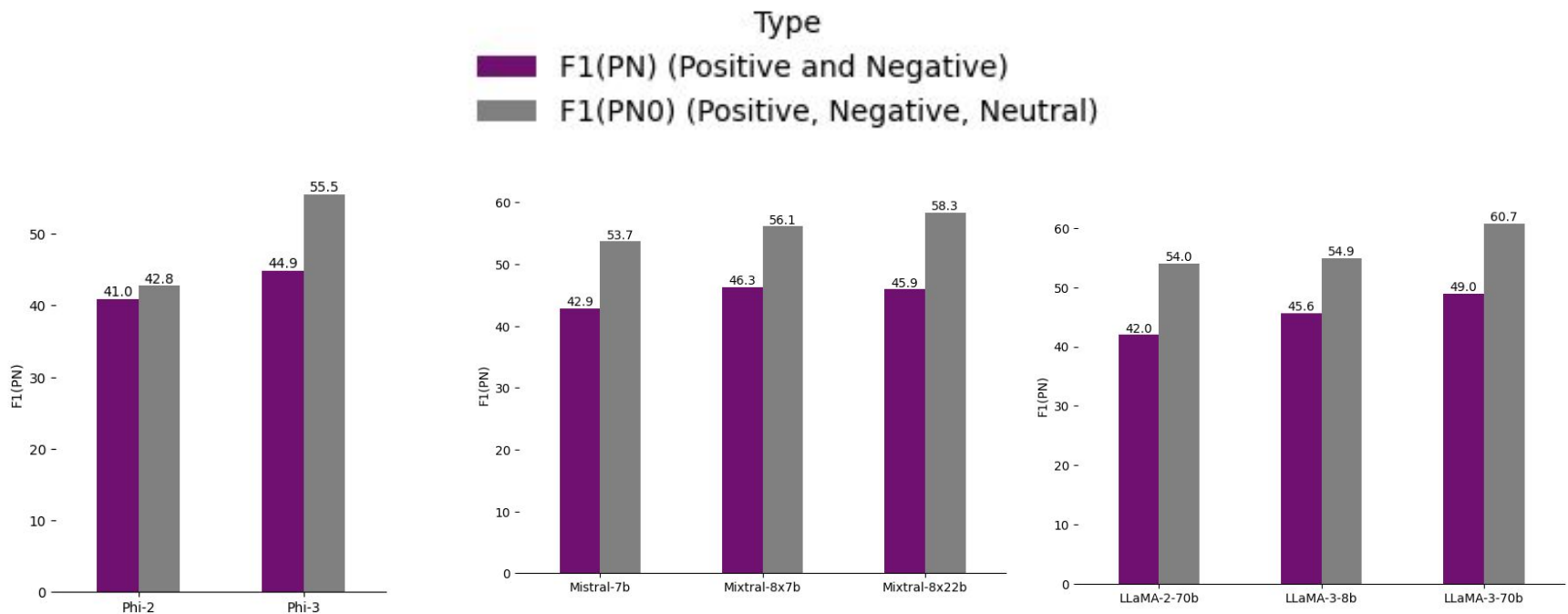- 1.8T   ChatGPT-4

🌡️ **temperature 0.1**

Findings 1/3 🔍 (Source Language Differences)

Findings 2/3 🔍 (**Open** LLMs Becoming better in factual presence of Sentiment)

Models are getting better in determining **factual presence of sentiment F1(PN0)**
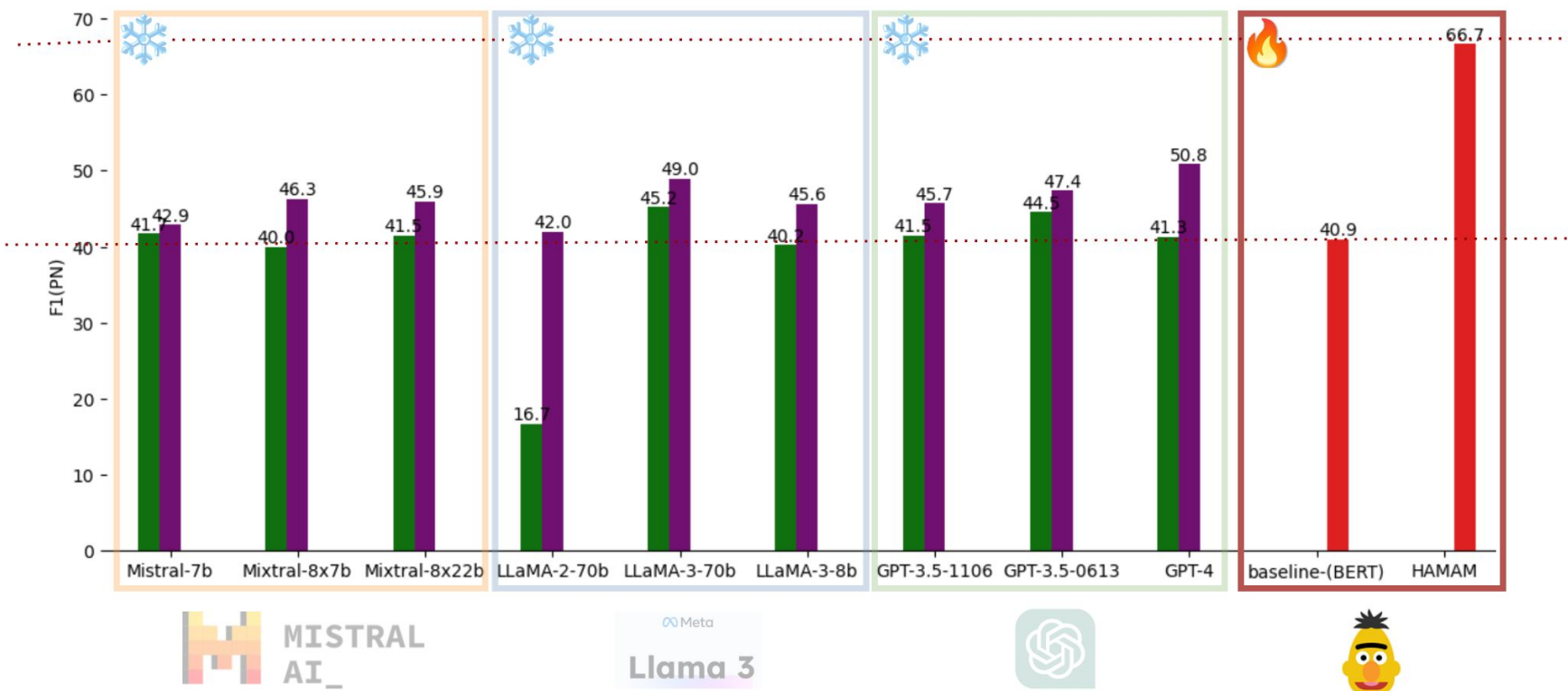
* Texts in English

Findings 3/3🔍 (Task goes beyond the LLM benchmarking)

Results could be significantly improved by fine-tuning

- Affection of bias
- Hallucinations
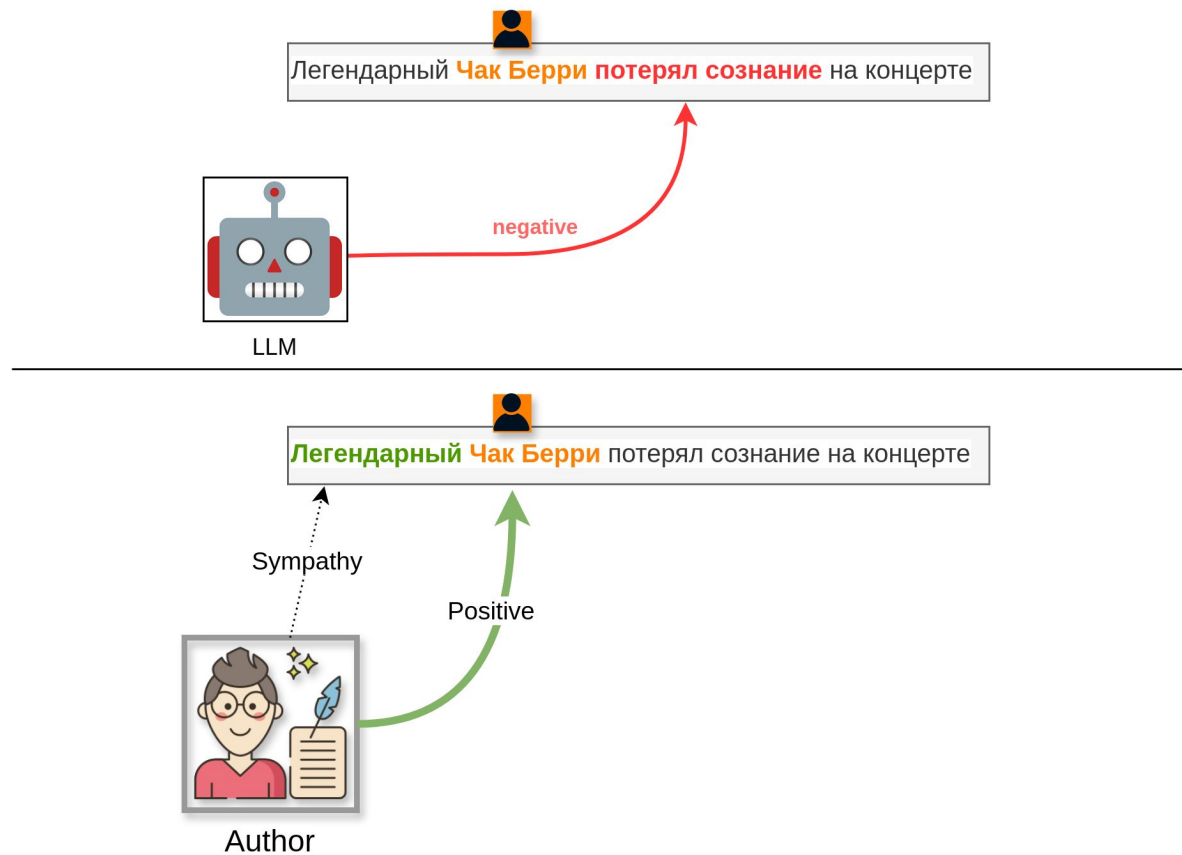
Best results in past by encoders:

- **BERT-base** (40.9)
- 🥇 **RoBERTa-large ensemble** (66.7)

Error Analysis 📝

Error Analysis 1/3 📝 (Sympathy missing)

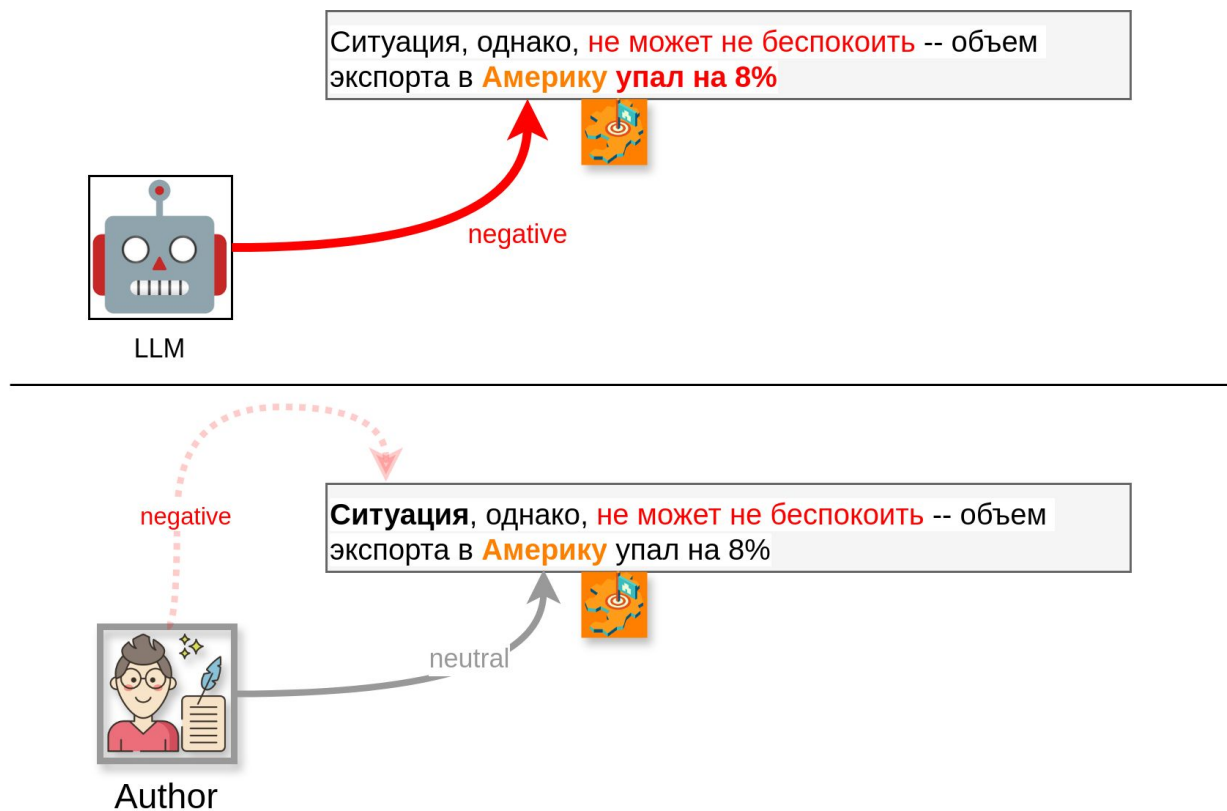*Легендарный **Чак Берри** потерял сознание на концерте*

Юлия *же в свою очередь обвиняет бывшего супруга в том, что он не выполняет решение суда …*
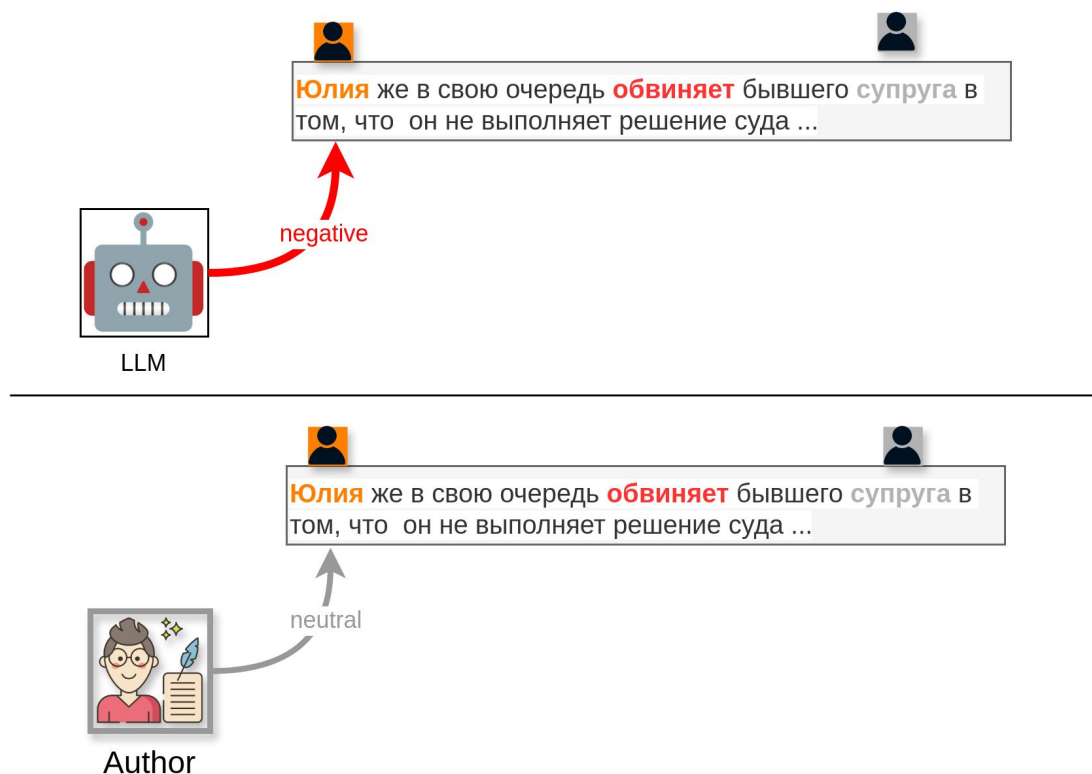
Ситуация, однако, не может не беспокоить -- объем экспорта в **Америку** упал на 8%

Reasoning in Steps 🔗

🔗 Zero-shot Chain-of-Thought (CoT)

💡 Generate the reasoning (supportive text) the answer

https://arxiv.org/abs/2205.11916

(c) Zero-shot

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: The answer (arabic numerals) is _____

(Output) 8 ✗

(d) Zero-shot-CoT (Ours)

Q: A juggler can juggle 16 balls. Half of the balls are golf balls, and half of the golf balls are blue. How many blue golf balls are there?
A: **Let's think step by step.** _____

(Output) *There are 16 balls in total. Half of the balls are golf balls. That means that there are 8 golf balls. Half of the golf balls are blue. That means that there are 4 blue golf balls.* ✓

Question: How this could be applied in Sentiment Analysis?

🔗 Fundamental Question of Rationality
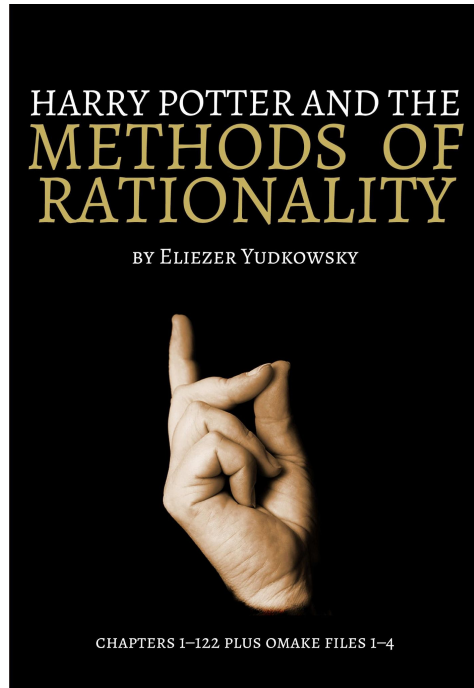
"I ask the fundamental question of rationality: why do you believe what you believe? What do you think you know and how do you think you know it? What makes you think Lucius wouldn't sacrifice you the same way he'd sacrifice anything else for power?"

Draco shot Harry another odd look. "Just what do *you* know about Father?"

"Um...seat on the Wizengamot, seat on Hogwarts' Board of Governors, incredibly wealthy, has the ear of Minister Fudge, has the confidence of

* 82 *

What and how if you breakdown the fundamental question of rationality.

- **Fact** (What do you think you know)
- **Opinion on it?** (how do you think you know it).

1 Given the sentence [SENTENCE].
Which specific <u>aspect</u> of **[ENTITY]** is possibly mentioned?

1 Given the sentence [SENTENCE]. Which specific aspect of **[ENTITY]** is possibly mentioned?

LLM

2 The mentioned aspect is [LLM Response]. Based on common sence what's the implicit opinion towards the mentioned aspect of **[ENTITY]** and why?

* Reasoning Implicit Sentiment with Chain-of-Thought Prompting
https://arxiv.org/abs/2305.11255

1

STEP 1. Given the sentence [SENTENCE]. Which specific aspect of **[ENTITY]** is possibly mentioned?

2

STEP 2. The mentioned aspect is [LLM Response]. Based on common sence what's the implicit opinion towards the mentioned aspect of **[ENTITY]** and why?

LLM

3

The opinion towards aspect of **[ENTITY]** is [LLM Response]. Based on such opinion, what is the sentiment polarity towards **[ENTITY]**?

4

Final inference for asking label → **Positive**, **negative**, neutral

LLM
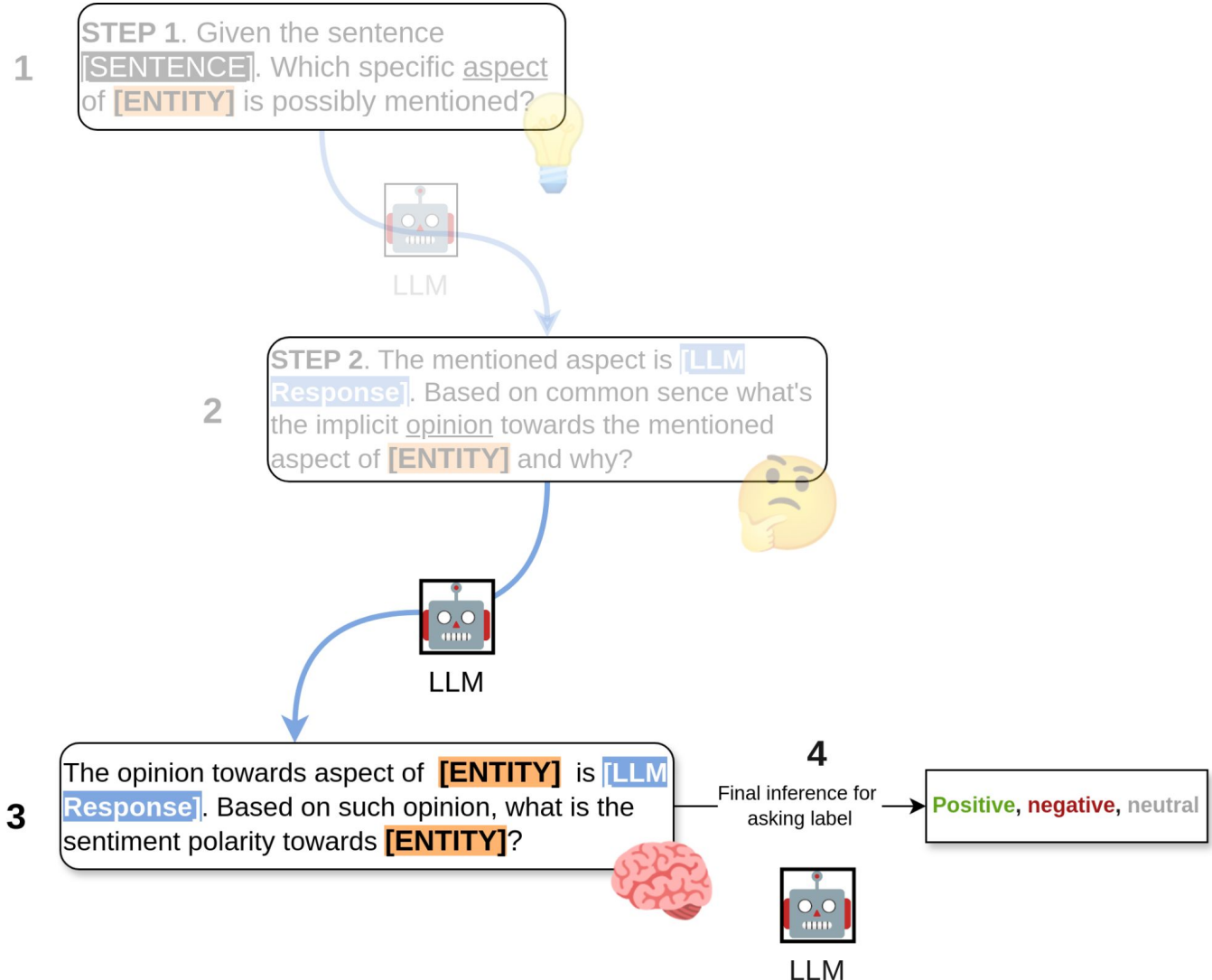
🔥 Fine-tuning with Chain-of-Thought

Experiments 🧪 (Fine-tuning)

Flan-T5: base (750M), large (1B), xl (3B)

Resource: NVIDIA-A100 (40GB)

Source: RuSentNE-2023 (train) in English

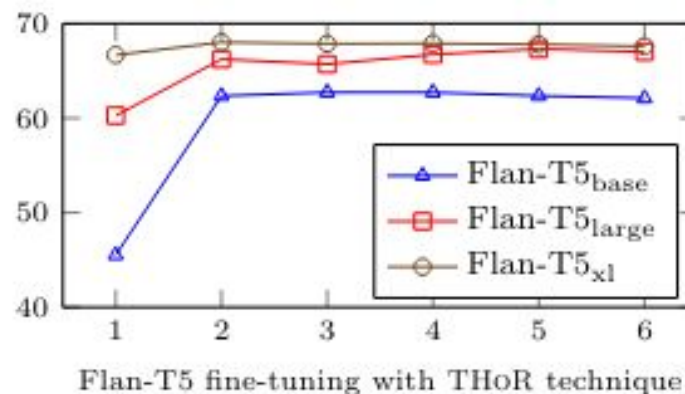**More details:** Reasoning-Framework launching in Colab

🐂**Training loop:**
1. Infer model per each CoT step (x3)
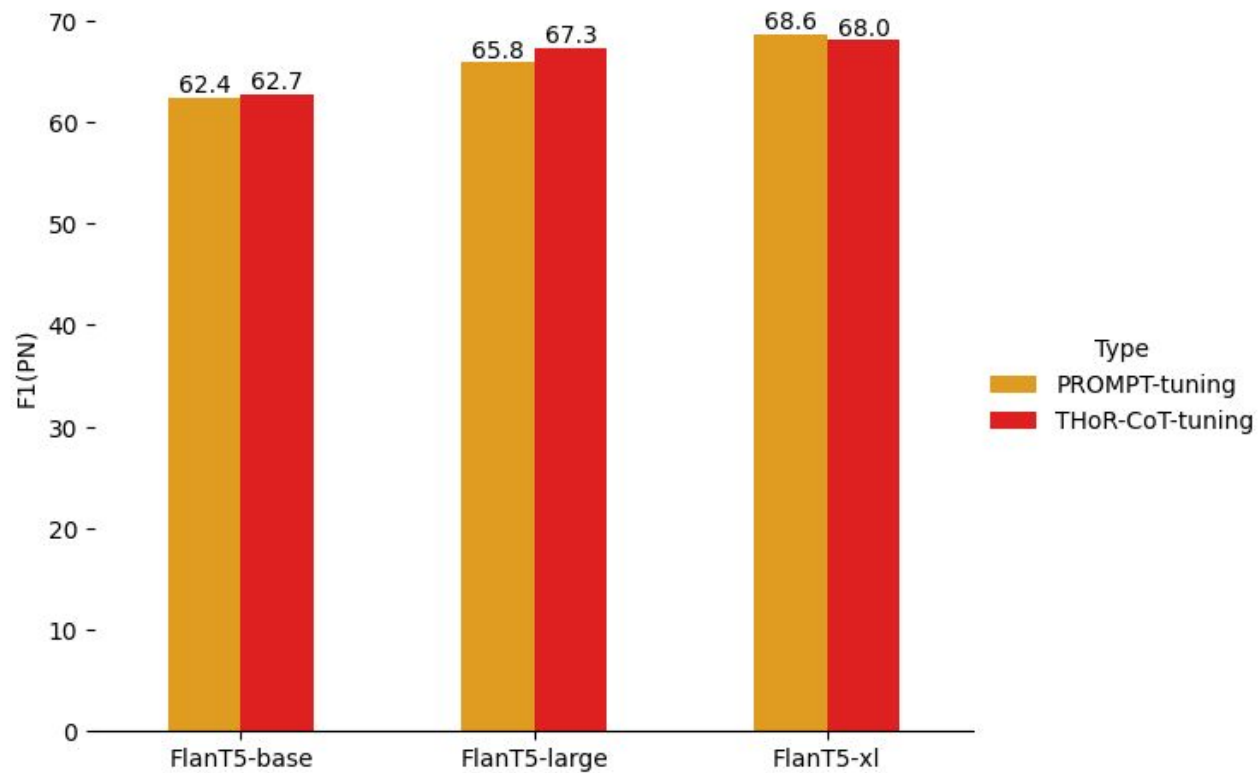2. Calculate *loss* and *backpropagation*

Answer checking strategy: label matching

🔥**Training duration**: *5-6 epochs.*
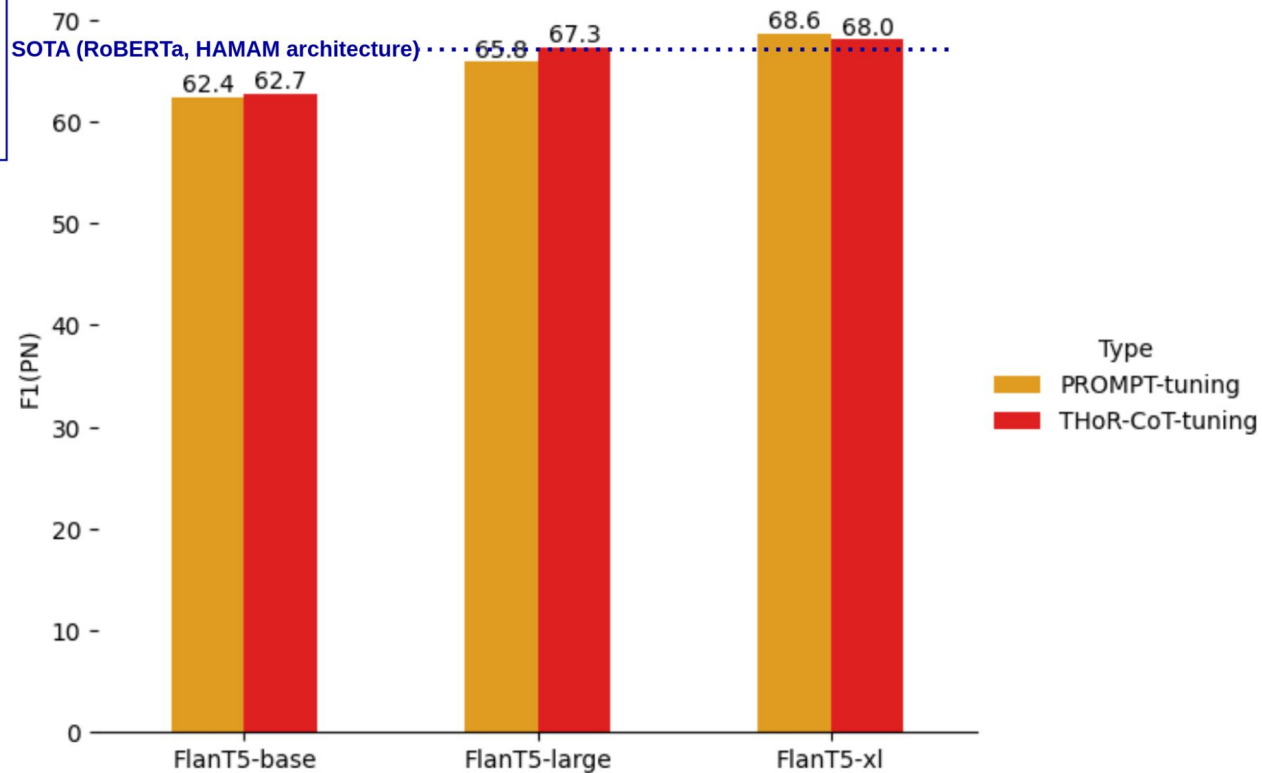

Flan-T5 fine-tuning with THoR technique

Results 📊 (Fine-tuning🔥)

CoT Fine-tuning help model robustness at the small scale

Results 📊 (Fine-tuning🔥)

And the result tuned FlanT5-xl **outperforms** the top encoder-based submission on **RuSentNE-2023**

Conclusions and Key Takeaways

1. 📊 How well **Large Language Models can actually reason** *as-it-is* depending on scales in Sentiment Analysis

    💡 Better to translate data

2. 🔗 How to apply **Chain-of-Thought** in Sentiment Analysis

    💡 Three Hop Reasoning Concept

3. 🔥 **How to build** the most-recent advanced Sentiment Analysis system

    💡 You can go with the THoR Tuning

    💡 larger model is preferable

Links

⌨️ **Framework:**

https://github.com/nicolay-r/Reasoning-for-Sentiment-Analysis-Framework

📰 **Paper:**

https://arxiv.org/abs/2404.12342

Leaderboard:

https://github.com/nicolay-r/RuSentNE-LLM-Benchmark

Thank you
for your attention!

Dr. Nicolay Rusnachenko

nicolay-r.github.io