Improving Generative Ad Text on Facebook using Reinforcement Learning

course: Natrual language processing

student name: hadi fathipour

student number: 40411334

Literature Review

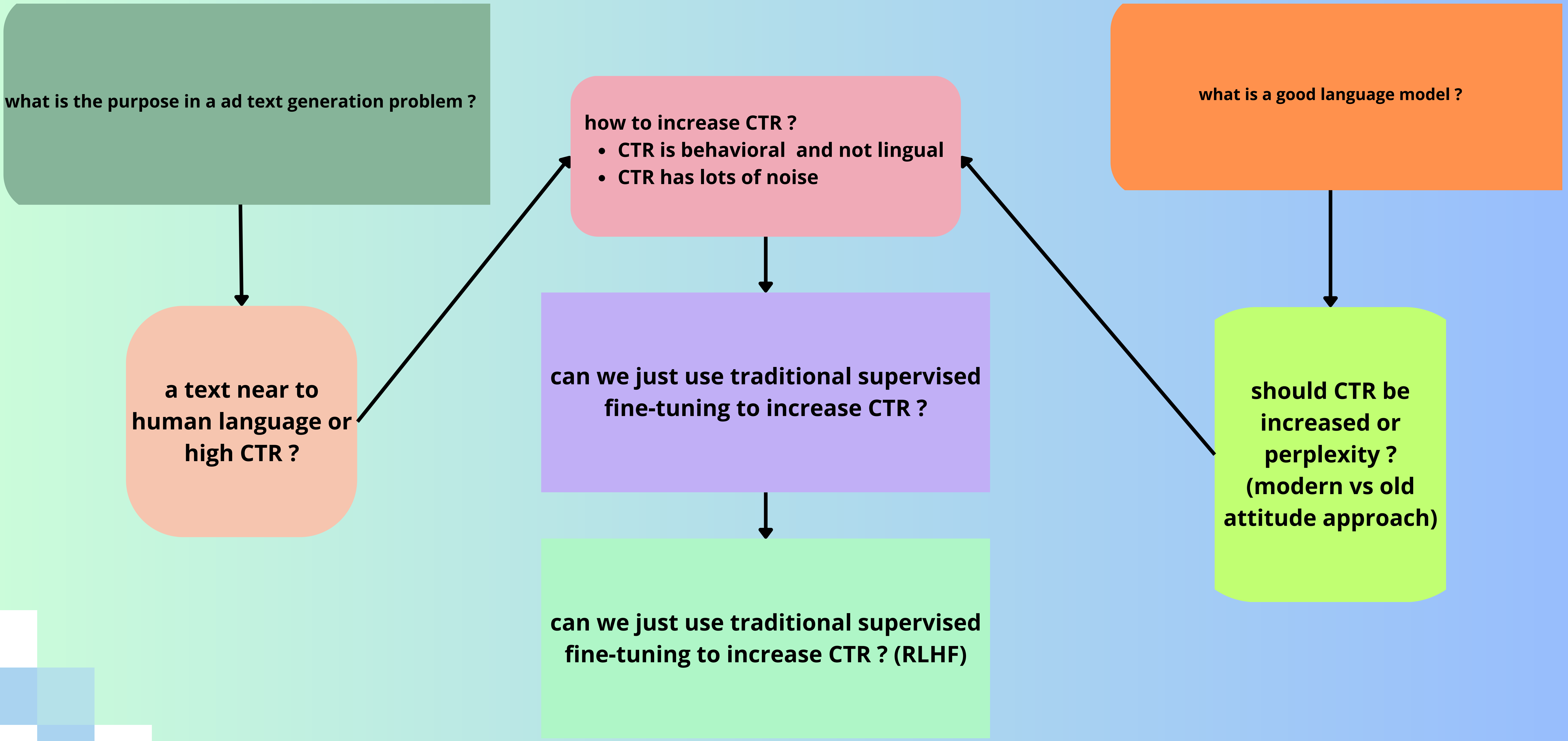
Overview of the Research Area

Recent advances in Natural Language Processing have shown that pre-trained large language models (LLMs) require an additional post-training or alignment phase in order to perform well on real-world tasks.

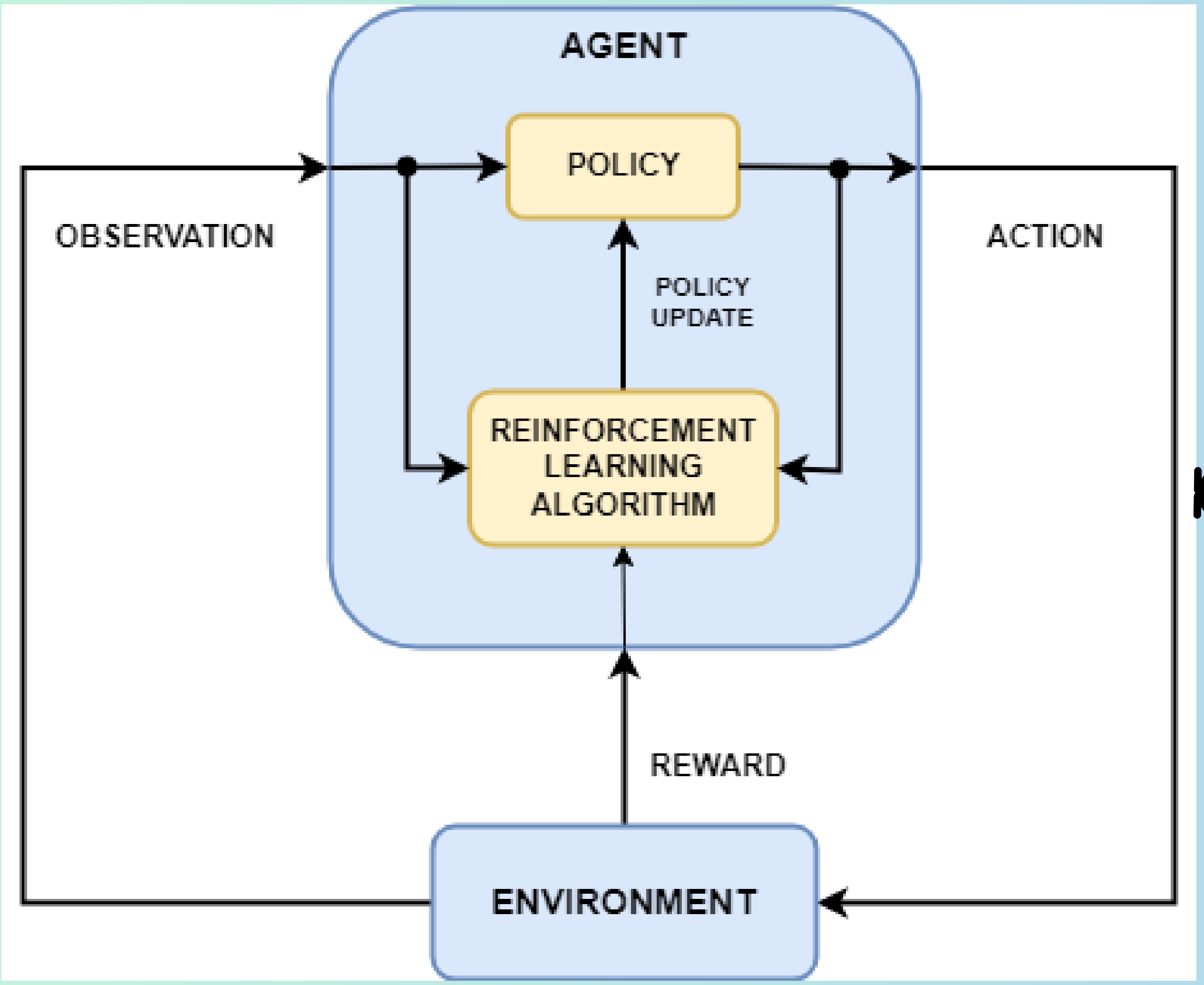
This line of research includes methods such as Supervised Fine-Tuning (SFT), Reinforcement Learning from Human Feedback (RLHF), and more recently, Reinforcement Learning using performance-based signals.

The common goal of these approaches is to align model outputs with downstream objectives that are not fully captured by likelihood-based training alone.

problem definition



solution: use reinforcement learning (RLPF)



reward:
CTR = num of clicks / impression

policy: LLM model
sth that estimates CTR for a given text (P(CTR|txt))

enviroment:
social media and user interaction

Paper 1 (Survey / Foundational Work)
Ouyang et al., 2022
“Training Language Models to Follow Instructions with Human Feedback”
Main idea:
This work introduces the RLHF framework, where human preference data is used to train a reward model, and the language model is further optimized using reinforcement learning (PPO).
Strengths:

- **Establishes a general alignment framework for LLMs**
- **Produces high-quality and controllable outputs**
- **Forms the foundation of systems such as ChatGPT**

Weaknesses:

- **Requires large amounts of costly human annotations**
- **Human feedback is subjective and noisy**
- **Limited scalability to domains with clear objective metrics**

Relevance:

- **This paper provides the theoretical and methodological foundation upon which later RL-based alignment methods are built.**

Paper 2 (Post-2022 Research Paper)
Deng et al., 2022
"RLPrompt: Optimizing Discrete Text Prompts with Reinforcement Learning"
Main idea:
This paper applies reinforcement learning to optimize discrete text prompts rather than model parameters, treating prompt selection as a policy optimization problem.
Strengths:

- **Demonstrates the effectiveness of RL for discrete text optimization**
- **Reduces reliance on labeled training data**

Weaknesses:

- **Focuses on prompt optimization, not full text generation**
- **Evaluated only in controlled experimental settings**

Relation to the main paper:
RLPrompt shows that reinforcement learning can successfully optimize text-based decisions, but it does not use real-world user behavior as feedback.

Paper 3 (Earlier Research on RL for Text Generation)
Shi et al., 2018
“Toward Diverse Text Generation with Inverse Reinforcement Learning”
Main idea:
This work uses Inverse Reinforcement Learning (IRL) to learn a reward function for text generation, with the goal of improving diversity and quality.
Strengths:

- **Early introduction of reinforcement learning into text generation**
- **Focuses on learning reward functions rather than hand-designed metrics**

Weaknesses:

- **Small-scale datasets**
- **No real-world or industrial evaluation**

Relevance:
This paper highlights the potential of RL-based approaches for text generation but lacks practical validation.

Paper 4 (Analytical / Critical Perspective)
Ramamurthy et al., 2022
"Is Reinforcement Learning (Not) for Natural Language Processing?"
Main idea:
This paper critically analyzes when reinforcement learning is suitable for NLP tasks and discusses challenges such as reward design, instability, and evaluation difficulties.
Strengths:

- **Provides a systematic analysis of RL challenges in NLP**
- **Offers a conceptual framework for understanding RL applicability**

Weaknesses:

- **Does not propose a large-scale applied solution**
- **Mainly theoretical and analytical**

Paper 5
Jiang et al., 2025
"Improving Generative Ad Text on Facebook using Reinforcement Learning"
Main contribution:
Introduces Reinforcement Learning with Performance Feedback (RLPF)
Uses real user behavior (CTR) as the reward signal
Evaluates the model through a large-scale A/B test in a real production system
Strengths:

- **Uses large-scale real-world data**
- **Demonstrates measurable business impact (+6.7% CTR)**
- **Represents one of the largest real-world evaluations of generative AI**

Limitations:

- **Uses offline reinforcement learning**
- **Optimizes a single metric (CTR)**
- **Does not explicitly model creativity or advertiser intent**

A/B Test real:
35,000 advertiser
10 week
640,000 ads
results:

- **+6.7% increase in CTR**
- **+18.5% in ad variation**

future works

- **Online Reinforcement Learning**
- **combination of CTR + Creativity**
- **the use of ad-provider in reward modification**
- **generalizable to :**
  - **Customer Support**
  - **Education**
  - **Public Messaging**

thanks for your attention 🌹