Reasoning in AI

Mordad 04

- In NeurIPS2019 conference Bengio delivered a keynote speech "from system1 deep learning to system2 deep learning "

Link: https://bdtechtalks.com/gninrael-peed-2019-spiruen-oigneb-auhsoy/2019/12/23

- This speech shaped a new field in artificial inteligence called System2

Link: https://youtu.be/Fw0pgyiKMBA

- Bengio asserted that current AI architecture are suffering from reasoning capabilities,he introduced his architecture called RIM

Link: https://arxiv.org/pdf/1909.10893v6

- The move number 37 showed that RL can learn patterns that human havent found yet

Link: https://www.youtube.com/watch?v=HT-UZkiOLv8

- RL can be used as RLHF and RLAIF

Link: https://arxiv.org/abs/2309.00267

- Many RL frameworks has been introduced for using it in the LM field like PPO,DPO,GRPO ans DQO

Link: https://arxiv.org/abs/2305.18290 https://arxiv.org/pdf/2402.03300 https://arxiv.org/pdf/2410.09302

- For mimicking human reasoning capabilities COT has been introduced

Link: https://arxiv.org/abs/2201.11903

To overcome the reasoning incapabilities four approaches of works has been done

- The zeroth approach tries to use merge connectionist AI and Symbolic AI

Link: https://journalwjarr.com/sites/default/files/fulltext_pdf/WJARR-.0287-2025pdf

- The first approach which Bengio has suggested is to develop a new architecture within internal reasoning capability
- The second approach is to guide the model in the inference time using PRM,ORM

Link: https://huggingface.co/docs/trl/main/en/prm_trainer

- One of the stunning works that have used this approach is MedSSS.

Link: https://github.com/pixas/MedSSS

- This work tries to guide the LM using a grpahreason

Link: https://arxiv.org/pdf/2308.09267v4

- The third approach is to train a language model using reasoning data an endeavour to mimic reasoning abilities.

- We have foucused on the last approach
- This work fine-tuned a small language model on rationale generated by large language model : https://arxiv.org/abs/2212.10071
- This work use DPO iteratively with a reward model to enhance reasoning ability : https://arxiv.org/html/2503.12854v1
- This work used DPO alongwith SFT to enhance COT reasoning : https://arxiv.org/pdf/2407.18248
- This work introduced learn by teacher paradigm: https://arxiv.org/abs/2406.14629
- this work introduced thinking policy optimization : https://arxiv.org/pdf/2410.10630
- this work tries to distill reasoning ability from a medical text book: https://arxiv.org/pdf/2404.00376

- This work tries to use encoding specificity hypthesis to enhance reasoning skills

Link: https://arxiv.org/pdf/2311.06985

- This work tries to enhance critic capabilities and use it for generating high quality reasoning

Link: https://arxiv.org/abs/2408.16326

- This work tries to determine the quality of rationale by reconstructing the question Link:https://arxiv.org/pdf/2305.11499
- This work tried to create a synthetic reasoning corpus and fine-tuned a LM to gain reasoning skills

Link: https://arxiv.org/pdf/2411.12498

- This work used MCTS to create high quality reasoning data

Link: https://arxiv.org/pdf/2501.01478