Assignment-4

Harshvardhan Patidar
Department of Artificial Intelligence
Indian Institute of Technology Hyderabad
`ai24btech11015@iith.ac.in`

1 Section 1.3

[2]

This section talks about the small advancements and new thoughts which slowly contributed to such successful modern AI tools. In the initial years (1943-1956) people had great expectations from AI. They were trying to model AI on the basis of conjecture that every problem can be so precisely described that a machine could be made to simulate it. Many successful programs like LT, GPS, Geometry Theorem Prover were also made. John McCarthy defined the Lisp language, which remained the dominant language for a long time. The concept of microworlds had also been employed to explore limited domains where AI could perform well. But these early systems failed on more difficult problems. The first reason being AI primarily based on informed introspection but not on a reliable algorithm. While the other reason was to overestimate the capabilities of trial and error method for solving problems.

Around the 1970s to 1980s, Early AI shifted from general-purpose methods to domain-specific expert systems like DENDRAL and MYCIN, which excelled in narrow fields. During the mid 1980s, back-propagation was revived causing great excitement. Learning from the limitation of the expert systems, AI adopted the use of probability, machine learning, and experimental results. Benchmark data sets became the standard to check the greatness of AI systems. These developments led to the reunification of AI with other subfields such as computer vision, robotics, speech recognition, multiagent systems, and natural language processing that had separated from AI long ago. Big data, or large data sets as a result of the World Wide Web motivated AI developers to develop an algorithm which takes advantage of this data to train AI models, which achieve remarkable accuracy. Banko and Brill remarked that improvement in performance obtained from increasing data size 2 to 3 folds is better than any improvement that can be obtained by tweaking the algorithm.

In 2011, Deep Learning, which is like multi-layered Machine Learning, has also seen success and hence started getting implemented in AI. Deep Learning relies on high computational power, big data sets, and better algorithms.

2 Section 1.5

This section introduces us to the benefits of AI, and also the risks associated with it. Looking at the positive side, AI has the potential to free human from repetitive work, increase productivity and efficiency, and also help in revolutionary scientific researches. But on the other hand, it can also prove to be a potential harm if used with wrong intention. Lethal autonomous weapons, persuasion of voters, biased decision making towards a specific community, impact on employment, cyber frauds are a few examples of misuse of AI.

The ideas of creating HLAI (Human level Artificial intelligence), AGI(artificial general intelligence) and ASI (Artificial Super intelligence) raises the concern that some day Human might need to face the "Gorilla Problem" because of the existence of AI, which might be taking up over the world, if goes out of control. "King Midas Problem" might be one other such problem that we may need to face. These issues can be tackled by making AI flexible and more aligned with human values along with the objective. Assisted games and Inverse Reinforcement Learning (IRL) are two methods being used in the ongoing effort to avoid the risks associated with ever improving AI.

3 AI Index Report

[1]

1. **Research and Development**: AI research is being dominated by industry rather than academic institutions. Total foundation models released got doubled in comparison to last year, along with the number of open-source models, US being the top source of these models. The cost of training models are skyrocketing, like Google's gemini ultra costed around $ 190 million. The AI patents have also increased substantially with maximum patents from China, hinting towards its probable upcoming dominance in the field. Open source projects in AI are also on a high-rise. Publications in the field of AI have also surged, reflecting the growing interest and activity in AI research.

2. **Technical Performance**: With improved algorithms, training and hardware, AI beats humans on some tasks like image classification and English understanding, but still fails in complex mathematics. To take benefit of this, multimodal AI are also being introduced, which may take input in multiple modes. The existing AI is being used to get better data, to train even better AI. Better means passing benchmarks, which are also got harder, like MMMu, MoCa etc. But this level of improvement in generative AI is increasing demand for human intervention in evaluation of AI. More technical research in agentic AI, which excel in specific environments has led to better agents, which are now being deployed in robots to make them more flexible.

3. **Responsible AI**: The Researchers are discovering complex vulnerabilities in LLMs. The reason being evaluations of AI are not standardized. Deepfakes are also another big threat. These are already affecting the elections throughout the world. AI developers are lacking transparency, especially regarding the training data. This is apparent from the fact that AI many-a-times output copyrighted materials from sources like The New York Times, which may invite legal battles. One such model, ChatGPT, is found to be politically biased, which reveals that its developer has supposedly used some political publications to train the model. AI is also difficult to analyze for extreme risks, as a result of which, the number of AI incidents continue to rise.

4. **Economy**: The investments in generative Ai is at an all time high, with US, strengthening its position as a leader in investment. Despite such high investments, The jobs in AI in the US and across the world are very few. A major chunk of organizations have increased their revenue and decreased costs by using generative AI. The industrial robot installation race has been led by China. Even more of the fortune 500 companies has started to talk about AI, a potential reason being AI helps in making workers more productive.

5. **Science and Medicine**: The AI is helping in accelerating scientific progress and field of medicine. GPT-4 Medprompt being a contributing factor. FDA is also approving more and more AI related medical devices.

6. **Education**: The number of bachelor's in CS continue to rise, but the number of Master's remain flat. But the number of Phds decrease as they are migrating towards industry. This is also true the other way around, transfer of talent from industry to academia has decreased. AI related degree programs are flourishing.

7. **Policy and Governance**: With AI's capabilities, the regulations against AI are also increasing sharply. In 2023, US policymakers made twice as many policies as in 2022. More and more regulatories agencies are now turning their attention towards AI, being aware of its misuse, trying to make it regulatable to avoid further misuse.

8. **Diversity**: In the U.S. and Canada the diversity among the CS graduates is increasing, with visible growth in the number of Asian, Hispanic, Black, and African American students. However, there is still a significant gender gap in european cs and informatics courses. The U.S. K–12 CS education is also becoming more diverse in gender and ethnicity, with rising participation from female and other minority students.

9. **Public Opinion**: People across the world are getting more cognizant of AI's potential impact and are getting more nervous about it. The Western countries are slowly getting more positive about AI products and services. ChatGPT is one of the most widely known and used model.

References

[1] The ai index 2024 annual report, 2024. `https://aiindex.stanford.edu/report/`.

[2] Stuart Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*. Pearson, New York, 4th edition, 2020.