Predicting the future is perilous

The Navy revealed the embryo of an electronic computer today that it expects will be able to walk, talk, see, write, reproduce itself and be conscious of its existence. . . . The service said it would . . . build the first of its Perceptron thinking machines that will be able to read and write. It is expected to be finished in about a year at a cost of $100,000.

– New York Times, July 8, 1958

Looking ahead

- Generating images, video, text, code, designing novel drugs and other chemicals – generative AI – will get more sophisticated.
- These create novel artifacts in in space of existing artifacts (hierarchical interpolation; neural network with relu provides hierarchical piecewise linear function).
- Constraining creations to be accurate is difficult.
- When used for high-stakes decisions, predictions will need to undergo critical evaluation to ensure they can be used as reliable components for the decision task.
- Better predictions can lead to better decisions if combined with appropriate values of stakeholders. Robust preference elicitation is needed.

Looking ahead (cont)

- Cases with no abundant data cannot learn from data alone – e.g., for most combinations of diseases no one has that combination; there is no data learn the interactions of these diseases.
- AI will be integral in science – creating and testing hypotheses. Predictions need to be empirically tested.
- An agent cannot learn from passive observation, such as text or video alone.
  - Text only contains what someone thought was interesting, and does not contain mundane details that everyone knows.
  - Video does not specify what would happen if agents involved did something different.
  - Embodied agents can play and carry out experiments in the would.

Future and Ethics of AI

- What will super-intelligent AI bring?
  - Automation and unemployment? What if humans are not needed to fulfill the needs of people (food, clothing, housing,...)
  - Smart weapons? Automated terrorists?
- What will a super-intelligent AI be able to do better?
  - predict the future
  - optimize (constrained optimization)
- Whose values/goals will they use? (Why?)
- Will we need a new ethics of AI designers and/or AI agents?
- Is super-human AI inevitable (wait till computers get faster)?
  Is there fundamental research to be done?
  Is it easy because humans are not as intelligent as we like to think?

"We can only see a short distance ahead, but we can see plenty there that needs to be done." – Turing 1950