Harnessing AI Conclusions

CS4000

WINTER 2024

Peter Denning
Producer

Our Objectives

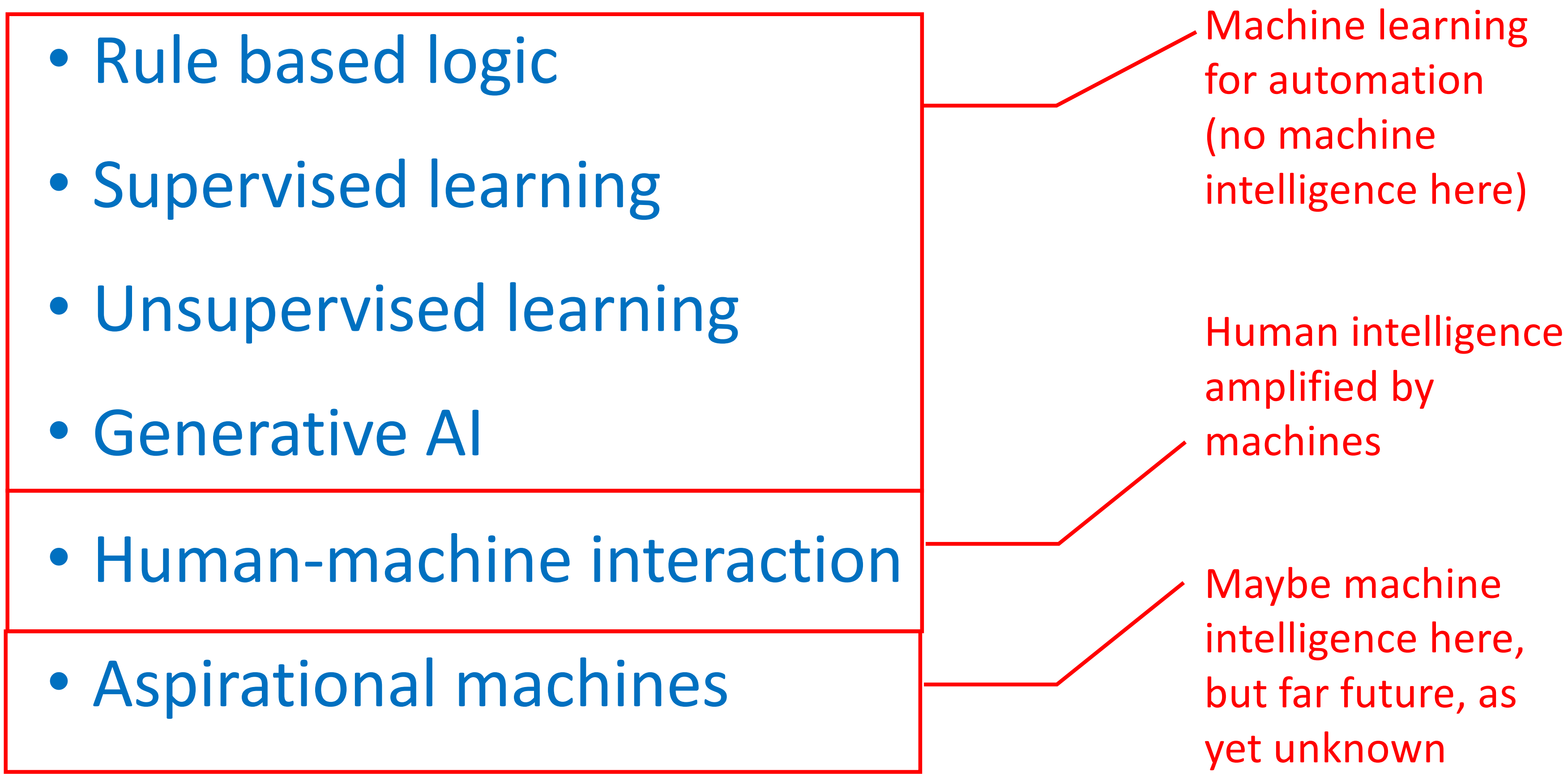
- Dimensions of AI
  - Machine learning hierarchy
  - Implementation in key domains
  - Strategy and risks
- Strengths and limitations
- Achieve reliability and responsible use
- The new face of battle
- See through the hype

Focal Points of Hype

- “AI Machines are intelligent or soon will be”
- “Large Language Models will become smarter than humans and eventually get out of control”
- “Dramatic breakthroughs in applications due to AI”
- “Programming is dead, long live machine learning”
- “Big data is the most important aspect of AI”
- “AI confers unbeatable geopolitical advantage”
- “The good far outweighs the bad”

CLAIM: “AI machines are intelligent”

Where is the machine intelligence?



- Rule based logic
- Supervised learning
- Unsupervised learning
- Generative AI

Machine learning for automation (no machine intelligence here)

- Human-machine interaction

Human intelligence amplified by machines

- Aspirational machines

Maybe machine intelligence here, but far future, as yet unknown

CLAIM: “LLMs will become smarter than humans and eventually get out of control”

- Smartness is very narrow on well defined tasks.
- LLMs are “parrots” that repeat back texts from training set that are statistically highly correlated with the prompt.
- The gap from LLM to general intelligence is enormous.
- Long list of things humans do in language that LLMs cannot do.

CLAIM: “Dramatic breakthroughs in applications due to AI”

- Data Science
- Management
- Vision
- Cyber security
- Natural language
- Robotics
- Ethics

Most of the progress comes from advances in computing machine power with special chips enabling the use of previously expensive AI algorithms and data analytics.

Some also comes from algorithm improvements.

CLAIM: “Programming is dead, long live machine learning”

- LLM codes are small; most software is much bigger
- Implementing AI is engineering with four components:
  - Designing and programming algorithms
  - Teaching from numerous examples
  - Reinforcement simulations
  - Rigorous testing
- Many complex, safety-critical software systems such as operating systems, simulators, and data management systems, require rigorous rule-based designs.

CLAIM: “Big data is the most important aspect of AI”

- Synergy: Big data and neural networks
- Require huge training sets
  - 300 billion words of text for GPT-4
  - 1 trillion parameters
  - Very expensive to obtain and train
  - Dubious quality training data
  - Energy intensive
- A lot of AI is programmed (e.g., training)
- Big data needed for ML but not all AI

CLAIM: “AI confers unbeatable geopolitical advantages”

- Speed
- Low cost of entry
- Ease of misuse
- Inability to regulate
- Adversarial attacks

CLAIM: “AI offers more good than evil”

- Don’t need Skynet for bad things to happen
- The good can be really good; the bad really bad
- AI powered ad-targeting poisoning social media for kids?
- Unable to verify claimed facts?
- Adversarial attacks render AI useless in war?
- Singularity and extinction of humankind?
- Slaughterbot scenario?
- Artificial stupidity?

Dilemmas

- Geopolitical consequences of AI
- Arms control and deterrence
- Fragility and intelligibility
- Hallucinations and fabrications
- Deep fakes
- Adversarial and trustworthy AI
- Bias
- High cost of reliable data
- IP and data gathering
- Human in the loop
- Civilian resistance to military uses of AI
- AI savvy workforce

We human operators need to find resolutions. The machines can support resolutions, but cannot find them.

Where to from here?

It is easy to become fascinated with the AI technology in and of itself. It is ingenious, clever, and cool.

But fascination with technology creates distance. A distance where others are pixels on a display. A distance that puts war on our screens as a video game, isolating us from its horrors and destruction. A distance that undermines our capacity for respect and increases our fear, anxiety, distrust, and resentment.

As you move forward after this course, be vigilant. Look for AI that helps you do your job better. Do not get lost in AI technology fascination. Do not lose sight of AI's weaknesses and limitations. Do not let an AI screen isolate you from understanding and respecting others.