NOTES from the reading:

I. The Shift in AI: From Rules to Learning

The landscape of AI changed significantly in March 2016 when Google DeepMind's **AlphaGo** defeated world champion Lee Sedol in the game Go. Turning point because:

- Unlike Deep Blue (chess), which followed human-programmed rules, AlphaGo used **machine learning** to develop its own strategies by analyzing millions of matches and playing against itself.
- Programmers create the algorithms and data sets, but they cannot always predict the specific moves the AI will make—the AI "learns by itself".
- Challenged the idea of human exceptionalism, leading to fears that machines might eventually outsmart and control us.

II. AI is Pervasive and "Invisible"

AI is now embedded in our daily lives:

- **Day-to-Day Tools**: It powers search engines (Google), targeted advertising (Facebook), and digital assistants (Siri/Alexa).
- **High-Stakes Systems**: AI is used in self-driving cars, autonomous weapons (drones), and even the legal system.
- Technology can now "read" human emotions via facial recognition and predict mental or bodily health without the user's knowledge.

III. Major Ethical & Societal Concerns

While AI offers benefits (like better cancer diagnoses) it introduces significant risks:

- **Algorithmic Bias**: Systems like COMPAS (used to predict criminal re-offending) have shown disproportionate "false positives" for Black individuals, reinforcing existing racial prejudices.
- **Privacy & Surveillance**: Facial recognition and predictive policing can lead to disproportionate targeting of specific socioeconomic or racial groups.
- **Information Integrity**: AI can spread hate speech (e.g., the Microsoft chatbot "Tay") or create "deepfakes"—false video speeches that look and sound like real people.

- **The Second Machine Age**: Machines are increasingly becoming **substitutes** for human labor rather than just complements, threatening to fundamentally change the social structure of work.

IV. The "Hype" vs. Reality

The discussion around AI is often split between practical ethics and "doom scenarios":

- **Superintelligence**: The idea that an "intelligence explosion" will lead to a **Singularity**—a point where machine intelligence surpasses all human intelligence combined.
- **Transhumanism**: The quest to use AI to "upgrade" humans, potentially achieving immortality or "Homo deus" (human-gods).
- **The Control Problem**: The worry that a superintelligent AI may not share human goals (e.g., the "paperclip maximizer" thought experiment, where an AI consumes Earth's resources just to make paperclips).
- **Critics**: Many experts argue that **General AI** (matching human intelligence) is decades away and that over-focusing on these sci-fi scenarios distracts us from the very real risks of currently deployed systems.

V. Cultural Narratives & "The Frankenstein Complex"

Our fears of AI are rooted in deep-seated Western narratives:

- **The Frankenstein Complex**: A term coined by Isaac Asimov to describe the fear that our creations (robots/AI) will inevitably turn on us, much like Mary Shelley’s monster or the Golem of Jewish legend.
- **Religious Parallels**: Concepts like the Singularity often mirror religious themes of "transcendence" (leaving the body behind) and "apocalypse" (the end of the world as we know it).
- **Alternative Views**: Other cultures, such as Japan (influenced by Shintoism), often view machines as "helpers" or friends rather than competitors, lacking the "Frankenstein" fear prevalent in the West.