**List of Concepts to Know**

Symbolic and Connectionist AI

Expert systems

Indian Chessboard

The hype cycle

Strong v weak AI

Artificial general intelligence (AGI)

The Turing Test

The Chinese Room

AlexNet

The dominance of deep learning

Moore’s Law

Input, hidden, output layers

Transfer learning

Long tail problem

Adversarial AI

Explainable AI

Deep fake

Natural language processing

IBM Watson

Transformers

Big Data

Data science

The singularity

Common sense

Dimensions of superintelligence

Paths to superintelligence

Solutionism

Intelligence Augmentation

Intelligent Infrastructure

Surveillance capitalism

Behavioral surplus

Extraction imperative

Prediction imperative

The “control revolution”

Economies of scope

Economies of action

Instrumentarian power

Moneyball

Weapons of Math Destruction

Mathematical models

Ground truth and proxies

Gamification

Elements of social engineering disasters

The gig economy

The takeover, complementarity, and bipolar distribution hypotheses The decomposition of professions

Tensions between professionals and clients

Problems with professional authority

Obstacles to the automation of professions

* **Symbolic and Connectionist AI:**  Two main approaches to AI. Symbolic AI involves manipulating symbols according to rules, while connectionist AI (or neural networks) is based on interconnected nodes that mimic the structure of the human brain.
  + **Ex (symbolic AI):** telling a computer exactly what to do: giving a robot a set of instructions on how to do something, and it follows those instructions step by step.
  + **Ex (connectionist AI):** computer learns from examples, just like when we learn from seeing things or playing games. Computer is shown a bunch of examples, and it figures things out by itself.
* **Expert Systems:** AI systems designed to emulate the decision-making abilities of a human expert in a particular domain by using rules and knowledge representations.
  + **Ex:** having a big book of rules – if you have fever, then it might tell you take medicine and rest. Those are the rules for treating fever.
* **Indian Chessboard:** used when talking about exponential growth, often applied to AI capabilities and technological progress. Chessboard start with one grain of rice, doubles each square. Closed problems can be unfathomably complex. We are bac at understanding exponential growth.
  + **Ex:** special checkerboard where you put grains of wheat, and each square gets twice as many grains as the square before it. Similarly with AI, it works in a similar way. Instead of wheat, we have a computer program that learns things. At first, it doesn’t know much, but as it learns it gets really smart (its knowledge continuously doubles, and at the end becomes a master.
* **The hype cycle:**  Model that describes the hype around new technology and how it progresses.
  + **Ex:** AI is like a new toy; everyone gets really excited about what AI can do thinking it will solve all our problems. This is the “peak of inflated expectations.” The people start to realize that AI isn’t as magical as they first thought, bringing it down to “the trough of disillusionment.” After a while, it scientists figure out how to use AI better; it starts to become useful. This is the “slope of enlightenment.” Then, it just becomes part of many things we do, not exciting but everyone knows its good to have.
* **Strong vs Weak AI:** Strong AI refers to systems that possess general intelligence that can meet or exceed that of humans. Weak AI refers to systems that are limited to the parameters they are programmed with, meaning they were designed only to do specific tasks.
  + **Ex (strong AI):** Strong AI is like a robot that can think and feel just like a person. It’s so smart that it can understand and learn from the world around it, just like you do.
  + **Ex (weak AI):** Weak AI is like a robot that can do specific tasks, but it doesn’t think or feel like a person. Imagine you have a toy robot that can only play music when you press a button. It can’t understand or feel anything; it just plays music.
* **Artificial General Intelligence (AGI):** AI systems that can meet or exceed human performance, very similar to human intelligence.
* **The Turing Test:** A test proposed by Alan Turing to determine a machine’s exhibit intelligent behavior indistinguishable from that of a human. In the test, a human evaluator interacts with both a machine and another human through a text interface and tries to determine which is which.
  + **Ex:** In class, when we had to figure out if it was the person answering our questions or ChatGPT. Surprisingly, it was very split in voting who said it.
* **The Chinese Room:** One person in closed off room from person outside of room. Person in the room does not know Chinese but has the tools to respond in Chinese. The outside of the room is fluent in Chinese, and will write in Chinese, and slip what they wrote, into the room with the person who doesn’t speak Chinese. Based on the characters that the person in the room sees, they will use their tools to respond, similar to how computers will run their programs without understanding their meanings. Developed by John Searle.
* **AlexNet:** A deep convolutional neural network architecture designed for image classification, which achieved breakthrough performance in the ImageNet Large scale visual recognition challenge in 2021.
  + **Ex:** A robot with special glasses (alexnet): these glasses help the robot to see and understand pictures. The robot with the glasses can be shown a picture of a cat and it can learn “Oh, that’s a cat!” After, if the robot is shown another picture of a cat, it’ll recognize that it’s also a cat.
* **The Dominance of Deep Learning:** the widespread adoption and success of deep learning algorithms, particularly neural networks with many layers, in various AI applications, including computer vision, natural language processing, and speech recognition.
  + **Ex:** convolutional neural networks: widely used for image recognition tasks, recurrent neural networks: designed to work with sequential data where the current output depends on previous inputs: natural language processing (language translation, text generation), time series analysis (stock prediction, weather forecasting)
* **Moore’s Law:** Observation by Gordon Moore, stating that machine computing power and capabilities will exponentially grow, thanks to the number of transistors on integrated circuits doubling every two years
* **Input, hidden, output layers:** The layers of nodes in a neural network. The input layer receives input data, the hidden layers process that data through weighted connection and activation functions, and the output layer produces that network’s predictions or outputs
* **Transfer learning:** A machine learning technique where knowledge gained from solving one problem is applied to a different but related problem, often leading to improved performance and efficiency. Deep learning models trained on specific task, training set and test sets usually same type of data
* **Long tail problem:** a phenomenon observed in distribution curves where many occurrences are clustered at the low end of the distribution (the “head), while a smaller number of occurrences extend far out to the right (the “tail”). In the context of AI, it refers to the challenge of addressing the diverse and less frequent cases or scenarios that are outside the mainstream. Deep learning models require tons of data, harder the task (for AI), the more data required. Deep learning models bad at recognizing what they don’t know.
* **Gradient Decent:** an optimization algorithm used in machine learning to minimize the error of a model by adjusting its parameters (weights) iteratively in the direction of steepest decent in the error of function
  + **Ex:** imagine you’re trying to find a hidden treasure in a large field, but you have no idea where it is. You start walking around randomly, taking a step and then looking around to see if you’re getting closer or farther from the treasure. **You keep doing this until you find it.** Basically, gradient decent it trying to minimize a mistake like trying to get as close to the right answer as possible. Start with a random guess and keep adjusting it a little bit at a time.
* **Adversarial AI:** generate counteract adversarial inputs. Exploit vulnerabilities. Changes data in ways that don’t matter to humans but change what AI sees.
* **Explainable AI:** Movement within AI to develop system that are open to human oversight, technical methods – decision trees, sparse linear models, and others. Sociological method human-in-loop-AI. AI systems and algorithms that can provide understandable explanations for their decisions and behavior, particularly in critical or sensitive applications where transparency is important.
  + **Ex:** In self-driving cars, it’s vital for engineers and users to understand why the AI system made certain driving decisions. Explainable AI helps improve safety and allows developers to identify and fix potential issues.
* **Deep Fake:** Media that uses AI assistance to mimic facial expressions, replace audio, video, or images of famous people/ popular media to create realistic but fake content.
  + **Ex:** using Taylor Swift’s voice in a commercial
* **Natural language processing:** The field of AI focused on enabling computers to understand, interpret, and generate human language, including speech recognition, language translation, and sentiment analysis.
  + **Ex:** the “brain” of a computer; basically helps say ChatGPT understand the context, grammar, and meaning behind the words I’m typing. NLP is the reason ChatGPT can converse with me.
* **IBM Watson:** AI (cognitive computing system) developed by IBM, famous for winning jeopardy in 2011. IBM thought Watson was the future and began selling it out to companies and hospitals, as Watson Health, and Watson was no more. Watson combines natural language processing, machine learning, and other AI techniques to analyze large volumes of data and answer questions posed in natural language.
* **Transformers:** A type of deep learning model architecture based on self-attention mechanisms, originally developed for natural language processing tasks but later applied to various other domains, achieving state-of-the-art results in many cases.
* **Winograd Schemas:** Faking out the AI. Tests the lingual abilities of AI, to understand and reason about natural language, particularly resolving ambiguous pronouns and references
* **Data:** Not a natural kind, information with a particular character, Information in a form suitable for processing by a computer. In the context of AI, data is crucial for training, testing, and improving machine learning models and algorithms.
* **Digital v.** **analogue:** Digital data is non-continuous, composed of discrete items, bits, pixels, zeros/ones, binary. Analogue is continuous, pictures, taped sound recordings, objects.
  + **Ex (digital):** Text documents: each character is represented by a specific code that the computer can understand and process
  + **Ex (analogue):** clocks: hands that move continuously around the dial represent the passage of time using a continuous motion
* **The data deluge:** The overwhelming volume, velocity, and variety of data being generated ad collected in the modern-world, presenting opportunities and challenges for AI and data-driven decision making
  + **Ex:** have box of toys that’s already super full, and then someone keeps adding toys to your box faster than you can play with them.Same goes with data, scientists get overwhelmed with a data deluge.
* **Big Data:** Extremely large datasets that are too complex to be processed using traditional data processing approaches. Big data often involves high-volume, high-velocity, and high-variety information sources. Not passive accumulation, but aggressive, active accumulation. Governments storing more surveillance data, businesses storing data on customer behavior
  + **Ex:** Walmart can keep track of what people buy, when they buy it, and even how they buy it. Every time someone buys something, Walmart’s computers take note of it. They collect all this information and use it to learn about their customers. So it a lot of people start buying umbrellas in a certain city because it’s raining, Walmart’s computers will notice and tell the stores in that city to put more umbrellas on the shelves.
* **Data Science:** interdisciplinary field that uses scientific methods, algorithms, and systems to extract knowledge and insights from structured and unstructured data.
* **The Singularity:** The idea that technology will progress far beyond human comprehension, leading to unprecedented changes in society.
* **Superintelligence:** intelligence that passes human intelligence in all forms. A potential outcome of advanced AI development
* **The intelligence explosion:** related to the singularity, the concept that once AI reaches a certain level of intelligence, it can rapidly self-improve and lead to a huge increase in intelligence.
* **Common sense:** basic level of practical knowledge and reasoning shared amongst humans with a culture or society. A challenging concept for AI as they can’t display or process emotion or understand moral contexts of things.
* **Dimensions of superintelligence:** Various aspects or capabilities that could characterize superintelligent AI, including cognitive abilities, creativity, emotional intelligence, and social understanding.
* **Solutionism:** Belief that everything can be solved with technology. Overly optimistic view of technology’s capabilities.
* **Surveillance Capitalism:** collection of personal data by companies to better understand our habits, preferences, and behaviors. In exchange, they offer free services like search engines, social media, and online shopping.
* **Behavioral Tech:** Information is gathered by corporations, that data is then embedded into technology to make predictions about our behavior, and to design systems that also shape our behavior. A reinforcing system that becomes more powerful with the more knowledge that is gained, better predictions from large amounts of knowledge/data gathered means a higher market demand for this technology.
* **Behavioral Surplus:** Behavioral tracking technology is everywhere, from apps to photo filters, data gathered from these things used to be considered useless as it served no purpose. Now these digital interactions are seen as opportunities to gather large amounts of personal data or in other words, a “Behavioral Surplus.” (more data than needed for reasons unrelated to the technological interaction)
* **Extraction Imperative:** Companies must collect as much data as possible on an increasingly massive scale.
* **Prediction Imperative:** The value of these companies is demonstrated in their ability to manipulate people.
* **The Control Revolution:** An approach to complex systems, involving information acquisition, processing, communication, and feedback. The dependence on information technology is rooted in the industrial revolution, which has led to revolution in societal control.
* **Economies of scope:** Production of one thing, reduces cost of another, increasingly demands economies of dept, not just location or social networks, but moods, personalities, fears
* **Economies of action:** Designed to manipulate/ modify behavior, more and more information is asked of you
* **Instrumentarian Power:** Uses of pools of aggregated user data to control human behavior techniques of prediction and manipulation
* **Moneyball:** Was first used in baseball, essentially looked at baseball from a statistical point of view using data, the theory was that more players on base would lead to more wins. Instead of relying on so-called “baseball experts”, teams started to rely on statisticians and mathematical models. This was essentially the start of mathematical models, being used across many different systems in society
* **Weapons of math destruction:** Mathematical algorithms that take on human traits and quantify them, which leads to damaging effects, and the perpetuation of bias against certain groups of people
* **Mathematical Models:** Models made to solve problems, using math and data, often speedy and efficient, and able to make predictions about outcomes
* **Ground Truth:** Any/all information that is known to be true or real, through direct observation, and empirical evidence
* **Proxy**: Not a directly related variable, but serves as a placeholder of an unobservable/immeasurable variable. Developed as stand-ins for real-world issues
* **Gamification**: applying games to help achieve tasks
* **Elements of social engineering disasters**
  + The administrative ordering of nature and society
    - State simplifications
  + A “high-modernist ideology”
    - An almost religious belief in the unerring progress of the sciences
    - Generalization from success of natural science and technology to social issues
  + An authoritarian state
    - Ability and willingness to use coercive power for social experiments
  + A weakened civil society
    - An inability to push back
* The gig economy: labor market that is centered around temporary/part-time work rather than full time careers
  + **Ex:**  uber/lyft
* **The Takeover Hypothesis (pessimistic take):** Technology will completely phase humans out of the workplace, leaving them with no jobs left to work
* **The complementary hypothesis (the optimist’s take):** technology will alter the labor market, and create new jobs rather than take them away, some will be eliminated but other will be made in that process, it may even create entirely new skills
* **The bipolar distribution hypothesis:** job losses from automation will be clustered in the middle skill/wage sector. These include factory jobs, repetitive white collar jobs, low skill jobs that require personal contact or involve highly irregular conditions, and jobs that are high skill and demand creativity and expertise. Surviving jobs will exist at the extremes
* **Decomposition of professions:** professionals do expert work, but with the rise in advanced technology, these professions may soon be broken down into decomposable tasks, rather than one long task.