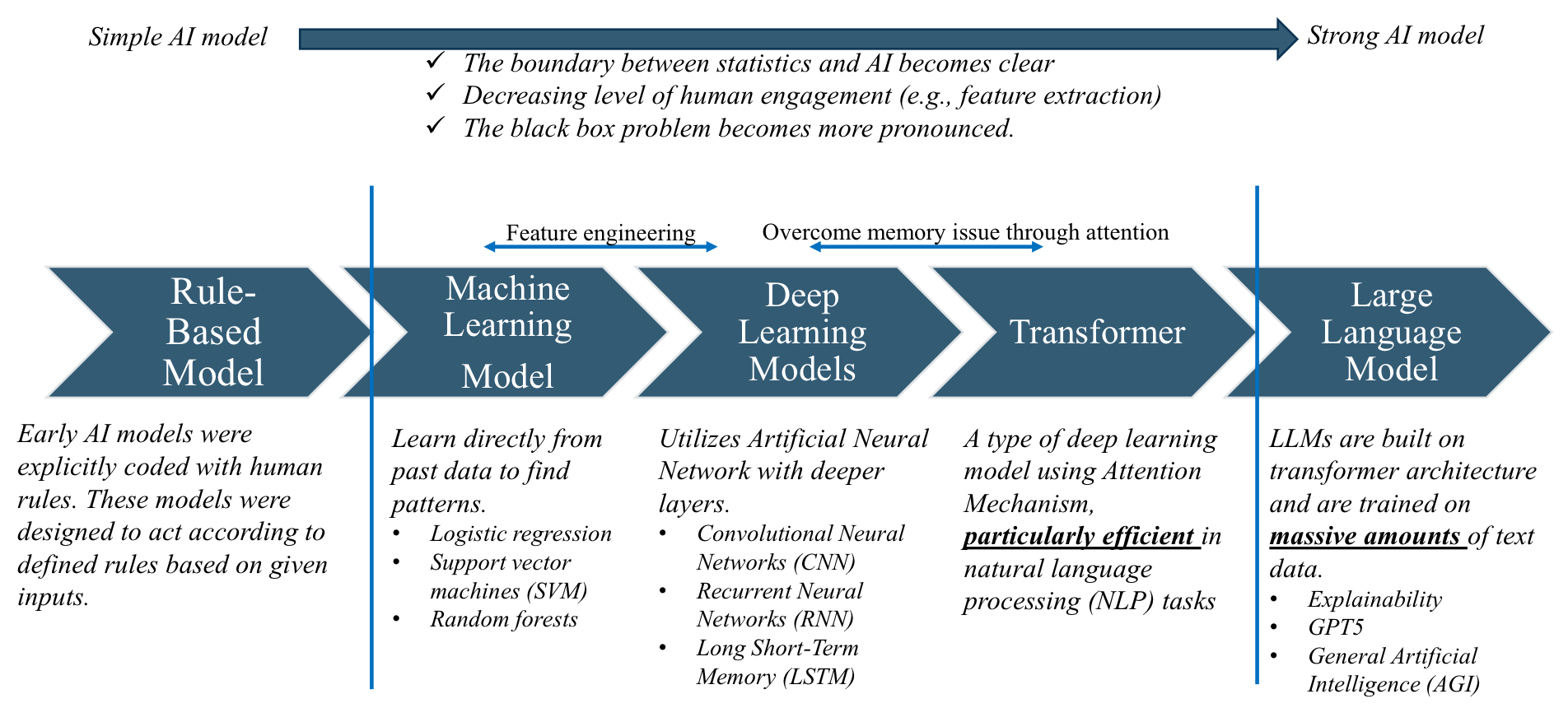
Private and public goods (H2 Econs stuffs)



HIERARCHY OF ALGORITHMS AND ARCHITECTURES

\*\* Meaningful data >>> A lot of data

Having the right model for the right task and utilising it in a way that adds real value

(smaller, more interpretable model is enough)

Deep learning prioritises **quantity of data**

* LLMs and transformers learn from vast amounts of data - more sophisticated and powerful language understanding

CNN - image and video

RNN - NLP

Given situation and context and resources

Intelligence

* comprehensive capability of an individual to reason about a problem and to solve it (cognitive and learning) LEARN + PROBLEM SOLVE

ARTIFICIAL intelligence

* simulation of human intelligence processes by machines, especially computer systems
* characterising AI is difficult because definition changes
* Goal: technology that ASSISTS humans

How to DEFINE AI - PERFECT artificial intelligence

* human-like
* mathematical reasoning
* common sense
* expert knowledge
* social behaviour
* rational
* achieve goals that have a performance measure
* Acting humanly - act sufficiently like a human to fool an interrogator
* Thinking humanly - comparing computer programs to human mind, observe human problem solving
* Thinking rationally - cannot separate AI vs human. Given correct premises, always give correct conclusions
* TURING TEST
  + prepare one interrogator and two responders (one computer and one human)
  + if interrogator cannot determine which is computer, computer passes the test

WEAK VS STRONG AI

Weak AI / narrow AI

* focuses on performing a specific task
* can only perform one type of task
* does not engage in conversation, emotion, learn for the sake of learning

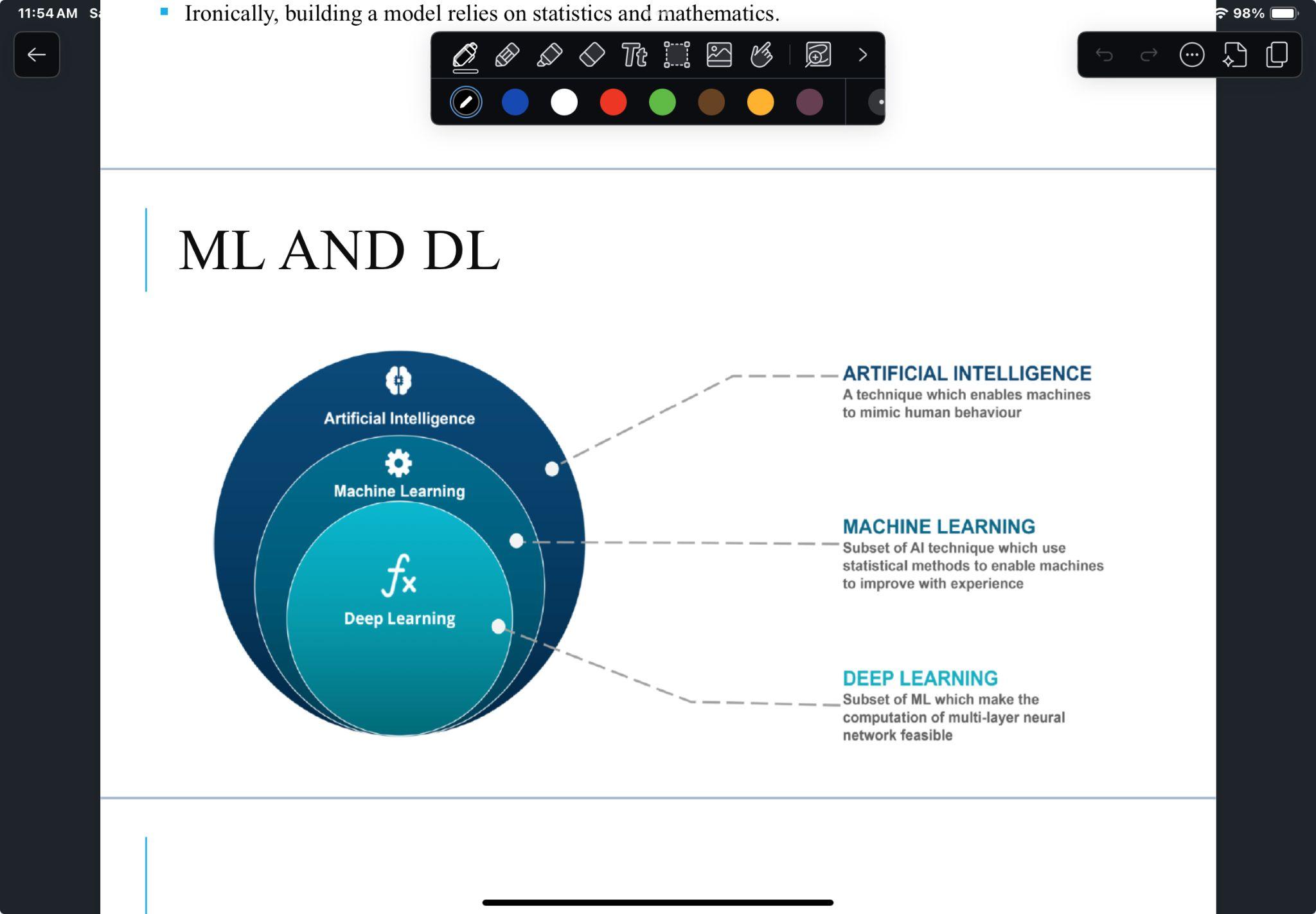
Strong AI

* teaching itself to solve new problems
* have emotions, purpose, humour

AI vs Statistics

* build models with the data instead of explicitly programmed instructions

ML and DL



PERCEPTION

The act of knowing or understanding something through our senses

* understand the meaning of sound + recognise situations that require comprehensive judgement skills
* contextual understand of conversation through NLP

PREDICTION - recommendation algorithm

* filling in missing information
* takes information you have and generate info that you don’t have - critical in current competitive market
* uses NLP
* e.g. Google email autocomplete, Netflix Personalised Video Ranking
* Prediction process
  + matching
  + find similar uses
  + compare preference between similar group
  + generate prediction results

Ethical implications

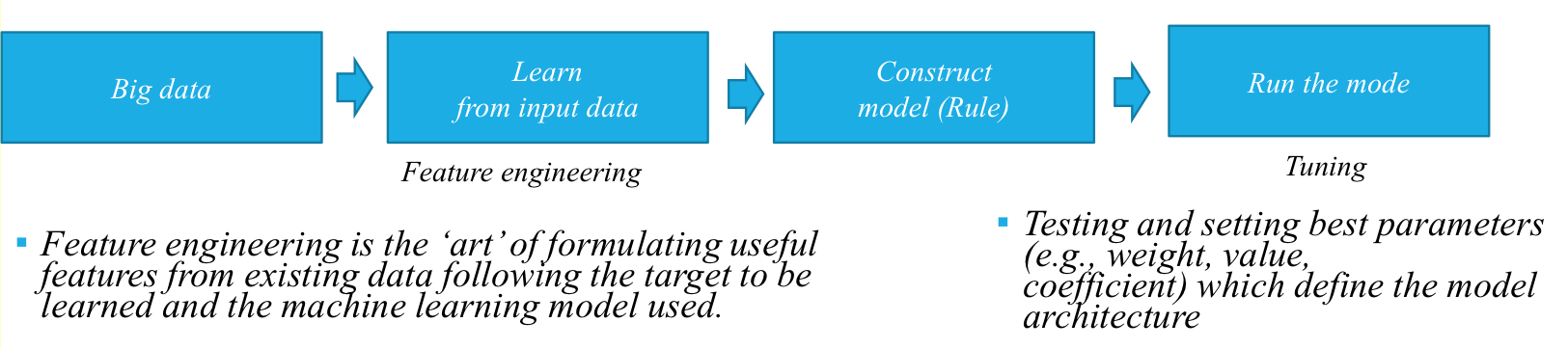
* human artists - why put in time to create anything when can be generated by AI
* become capable of non task oriented conversations
* questionable whether AI can fully replace human due to psychological resistance

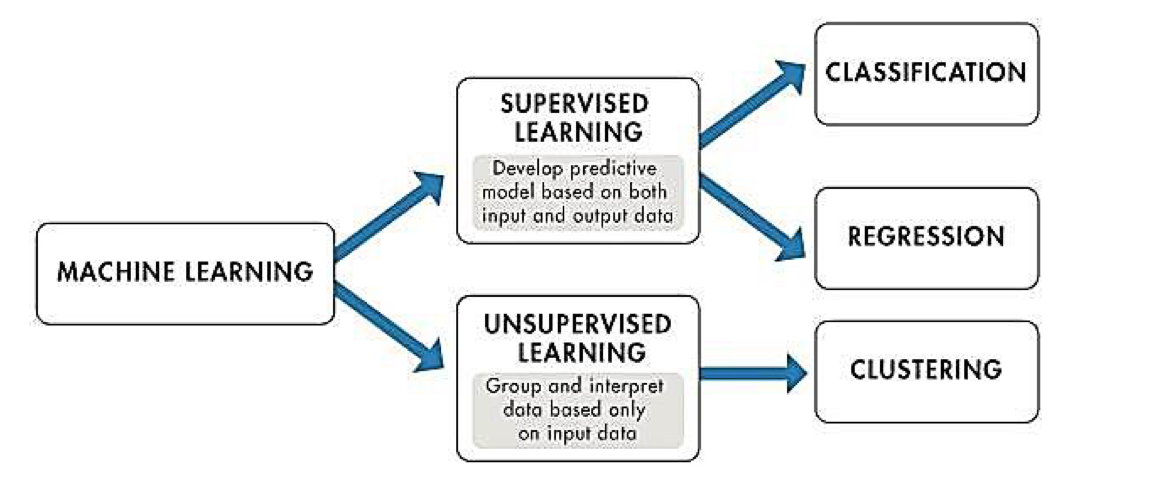
GENERATIVE AI

Branch of AI that focuses on creating new content (NOT recognising patterns or making predictions)

* (LLM) model is fed huge datasets and learns to predict the next word in a sentence based on context of words before it
* iterative process, does not yield perfect results
  + refine prompts, review outputs, gradually improve accuracy and quality of results through continuous interaction
* AI automation allow machines to handle complex interactions that require human judgement

MACHINE LEARNING - WHAT IS?

* Find and learn the rules by itself by training using large-scale data
* Obtain meaningful results with a few hundred to few thousand data points
* Based on statistical methods/math
* 
* Automates analytical model building through big data - unable to explain the logic/reason behind the output
* Aims for accuracy rather than mathematical rigour
  + goal is to reduce the error between the predicted value and the actual correct answer



**SUPERVISED LEARNING**: use of **LABELLED DATASETS**

* datasets are designed to train or supervise algorithms into classifying data or predicting outcomes
* model can measure its accuracy and learn over time
  + classification
  + regression
* labelled datasets are expensive and requires effort
* lack of responsiveness to new data
* data is very well known and labelled
* supervised method are more accurate and reliable

**UNSUPERVISED LEARNING**: use of machine learning algorithms to analyse and cluster **UNLABELLED DATASETS**

* discover hidden patterns in data without *human intervention*
* work on their own to discover the inherent structure of unlabelled data
  + still require human intervention for validating output variables otherwise wildly inaccurate results
  + clustering -> grouping unlabelled data based on their similarities / differences
* inaccurate in many cases, computationally complex (dog vs muffin example)
* need a large training set to produce intended outcomes

**SEMI-SUPERVISED LEARNING** contributes a **small amount of labelled data** with **large amount of unlabelled data** during training / weak supervision

* reduce data acquisition costs
* improve learning accuracy
* generate pseudo labels -> use only the most confident dataset

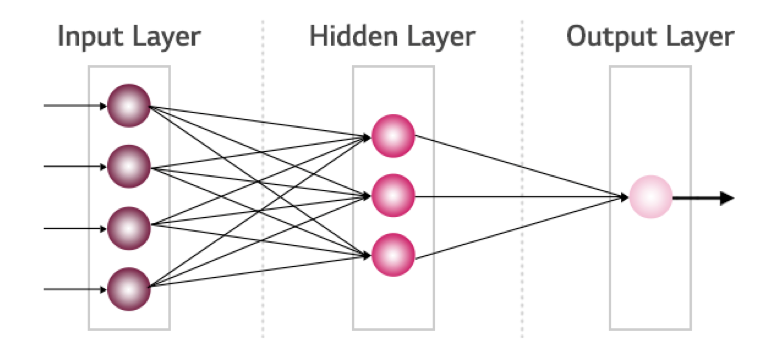
**REINFORCEMENT LEARNING** based on rewarding desired behaviours / punishing undesired ones

* perceive and interpret its environment
* take actions and learn through trial and error
* e.g. AlphaGo
* uses reward data from action of agent in given environment

DEEP LEARNING

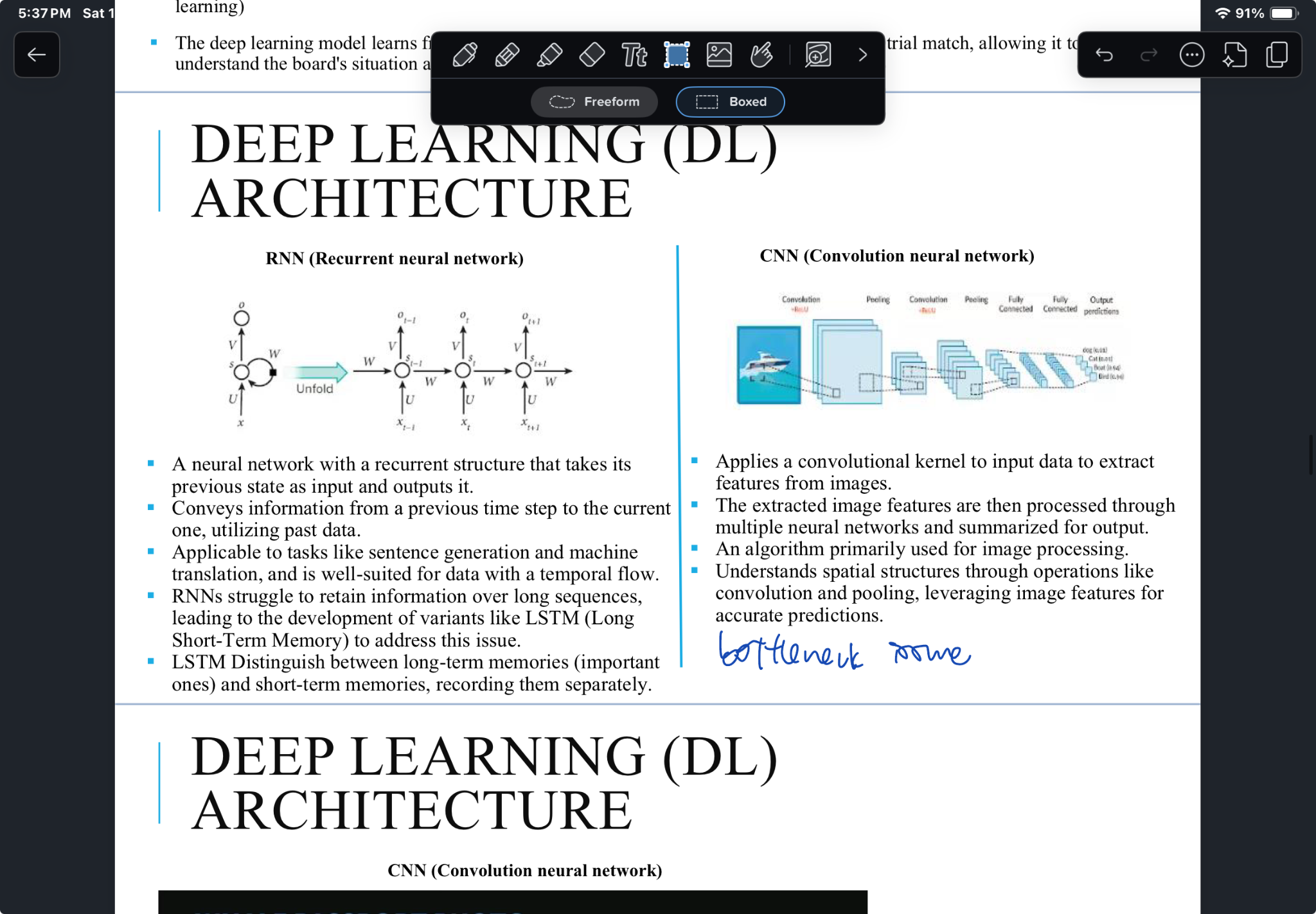
Uses a sufficiently deep ANN -> DNN (deep neural network)

Issue of still being a black box

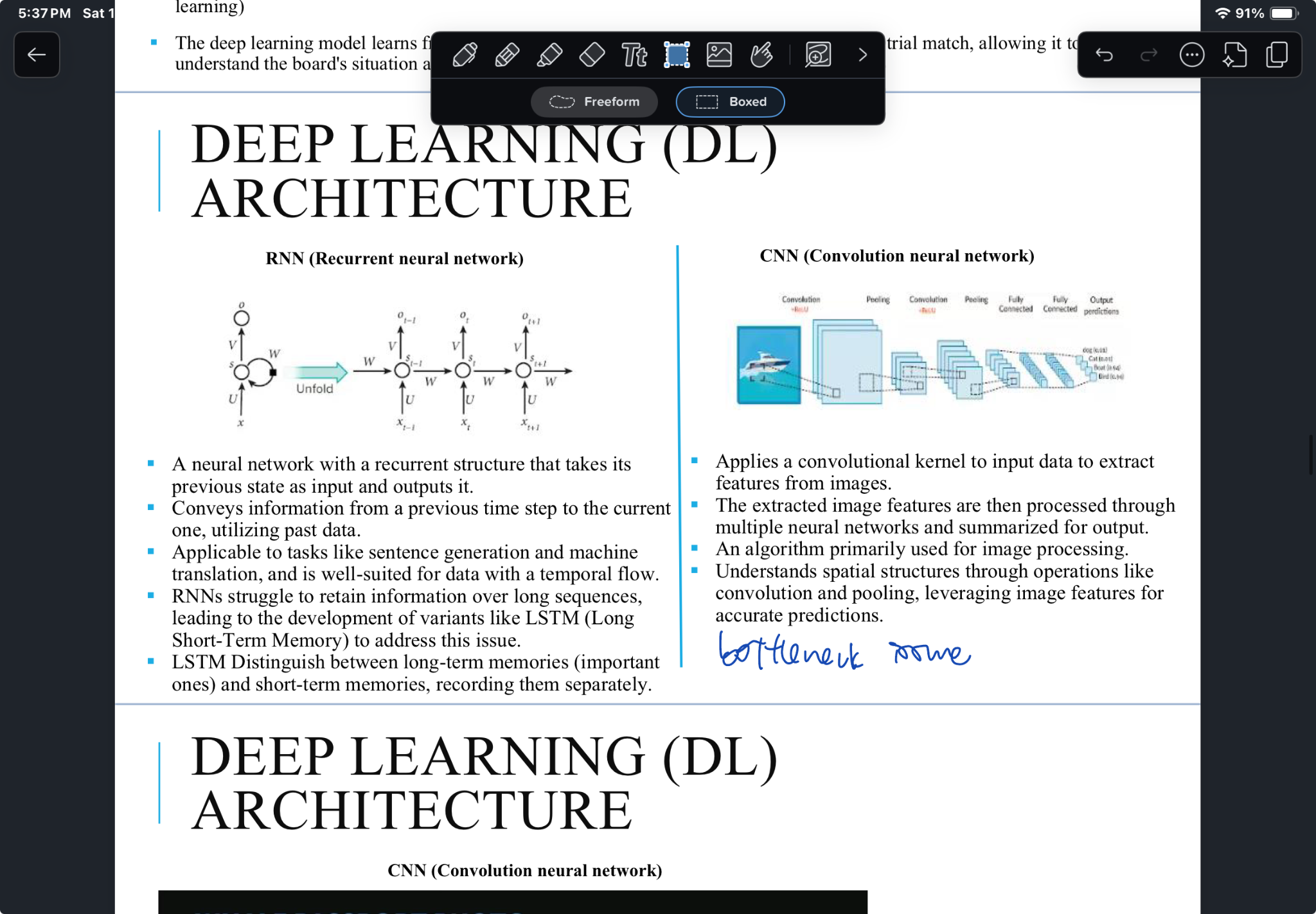
* does not process the data all at once
* divides into multiple parts for training
* **batch size**: size of the data chunk processed at one time
  + input layer - initial data for NN
  + hidden layers - perform transformations of inputs using weights and activation functions
  + output layer - produce result
  + 

**DEEP LEARNING ARCHITECTURE**

**RECURRENT NEURAL NETWORK (RNN)**

* used in Natural Language Processing
* Recurrent structure that takes its previous state as input and outputs it
* utilise past data by conveying information from a previous step to the current one
* sentence generation, machine translation (*data with a temporal flow*)
* struggle to retain information over long sequences
* LSTM -> record long term and short term memories separately
* context vector?
* ****

**CONVOLUTIONAL NEURAL NETWORK**

* used in image and video processing
* applies a convolutional kernel to input data to extract features from images
* processed through multiple neural networks and summarised for output
* understands spatial structures -> convolution and polling
* bottleneck issue
* 

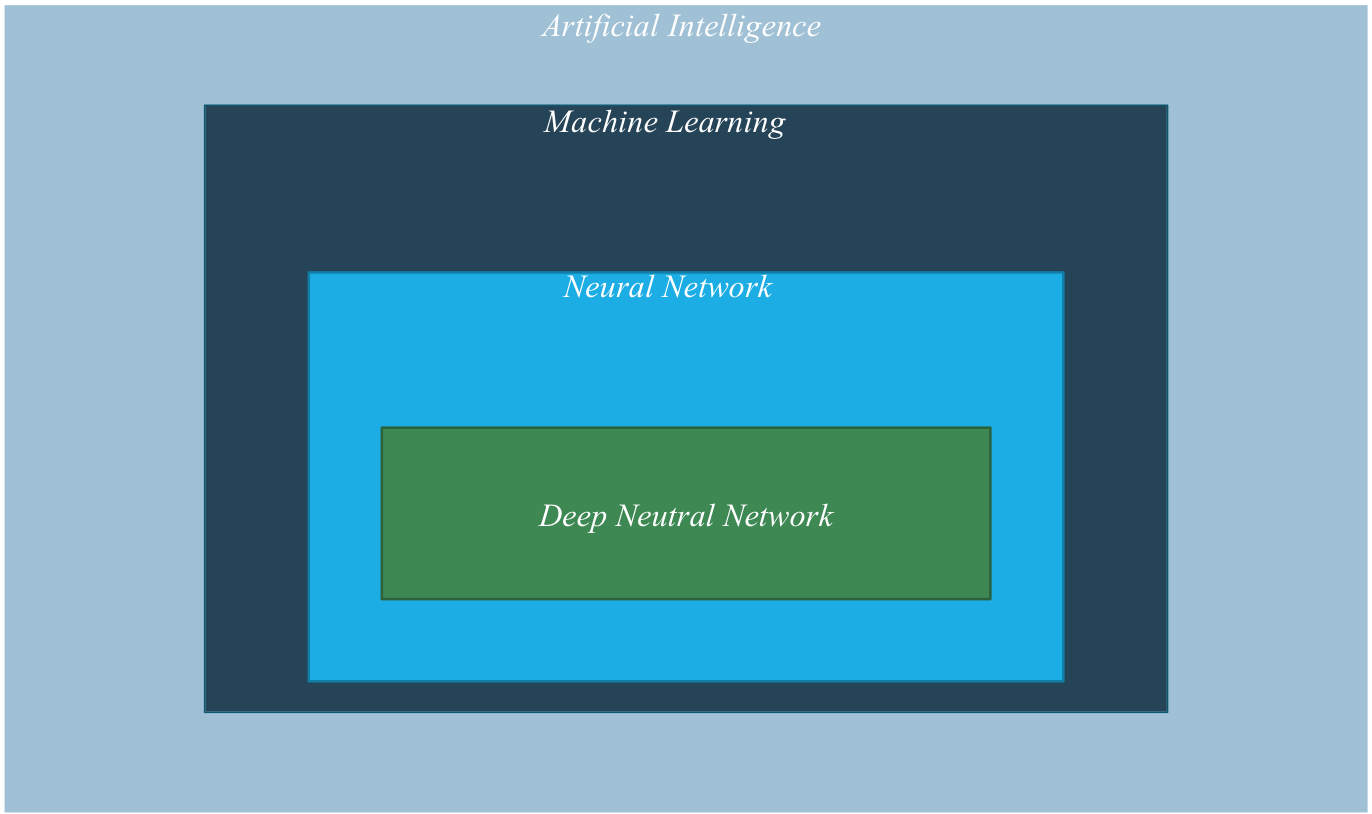
Advantages

* eliminate need for data labelling
* optimisation
* high accuracy

Disadvantages

* require large amounts of data
* requires computing ability and data storage
* **black box issue**: impossible to look inside to see how it works, reasoning of neural network is embedded in behaviour of thousands of simulated neurons

Fun visualisation



LEADING AI LANDSCAPE - GOOGLE

AI first, then AI for everyone

Google AI Strategy

Google Cloud -> cost effectiveness of AI

Development tools and API - engineering and programming effectiveness

Lecture 11 notes based on Notability AI Learn

​​

## Machine Learning (ML) Fundamentals

- ML systems learn rules autonomously through training on large-scale data

- Requires minimum 1,000 data points to avoid overfitting

- Uses statistical methods and mathematics for rule generation

- Process includes feature engineering, model construction, and parameter tuning

- Focuses on accuracy over mathematical rigor

- Often cannot explain logic behind outputs

### Types of Machine Learning

#### Supervised Learning

- Uses labeled datasets for training

- More accurate and reliable results

- Applications include classification and regression

- Challenges:

- Expensive to obtain labeled datasets

- Limited responsiveness to new data

- Data security concerns

#### Unsupervised Learning

- Works with unlabeled data to discover patterns

- Requires less effort than supervised learning

- Uses clustering techniques

- Limitations:

- Needs human validation

- Computationally complex

- Requires large training sets

#### Semi-supervised Learning

- Combines small amount of labeled data with large unlabeled datasets

- Benefits:

- Reduces data acquisition costs

- Improves learning accuracy

- Generates pseudo-labels for new data

#### Reinforcement Learning

- Based on rewarding desired behaviors and punishing undesired ones

- Learns through trial and error

- Example: AlphaGo learning through historical games and self-play

## Neural Networks and Deep Learning

### Neural Networks

- Artificial neural networks (ANN) mimic human brain structure

- Approximates cognitive processes using mathematical representations

- Uses weights and vectors for computation

### Deep Learning

- Uses deep neural networks with multiple layers

- Components:

- Input layers: initial data entry

- Hidden layers: perform complex transformations

- Output layers: produce results

- Critical parameters:

- Activation functions (ReLU, Sigmoid, Tanh)

- Loss functions

- Optimization algorithms

- Batch size for processing

### Deep Learning Architectures

#### CNN (Convolutional Neural Network)

- Specialized for image processing

- Uses convolutional kernels to extract features

- Processes spatial structures through convolution and pooling

#### RNN (Recurrent Neural Network)

- Designed for sequential data

- Uses previous states as inputs

- Applications: sentence generation, machine translation

- Variants like LSTM handle long-term dependencies

### Advantages and Disadvantages of Deep Learning

Advantages:

- Eliminates need for data labeling

- Provides high accuracy and optimization

Disadvantages:

- Requires large amounts of data

- Needs significant computing power

- Black box issue: difficult to understand internal decision-making

## Business Application: Google Case Study

- Implements AI First strategy

- Focus on AI For Everyone initiative

- Key approaches:

- Open innovation ecosystem

- Cloud platform development

- Open-source libraries and APIs

- Strategic acquisitions and expert recruitment

- Practical applications:

- AR navigation using image recognition

- Global Fishing Watch for detecting illegal fishing

- Real-time data sharing and analysis

TRANSFORMER architecture

**Context Vector**

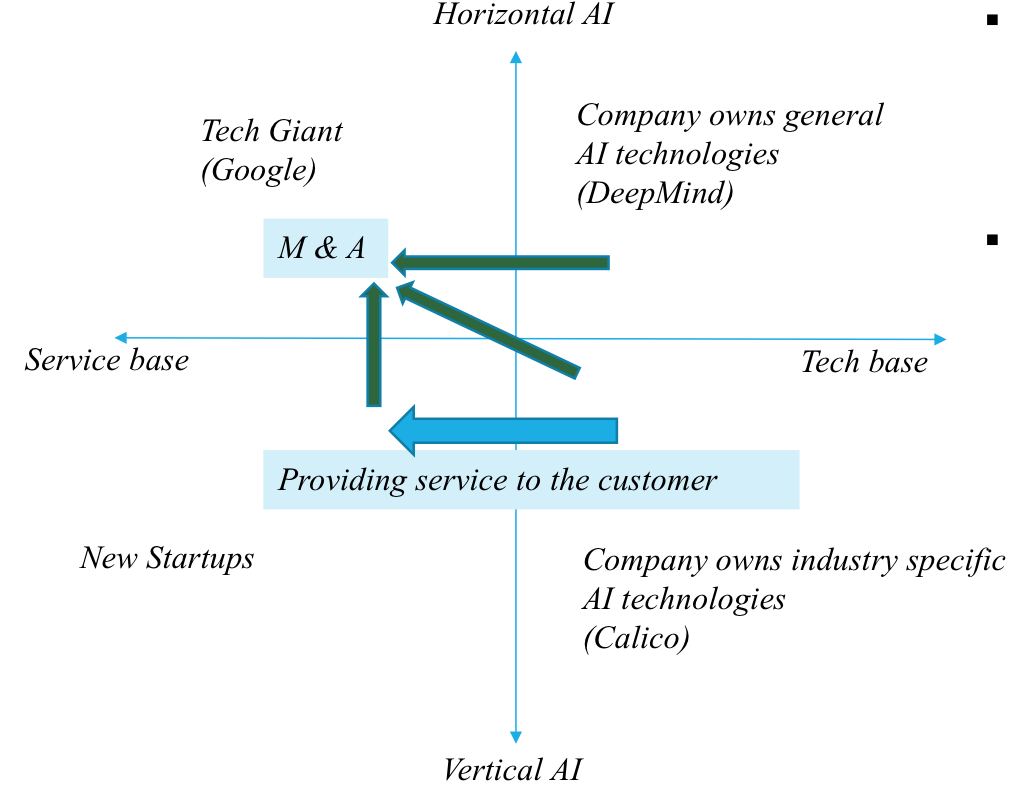
* fixed length vector comes from the last state of the encoder
* represent a compressed form of the entire source sentence
* used by decoder to aid in generating the target sentence
* limited in capturing all the information of source sentence
* (unable to) encompass all its complexities and nuances
* *bottleneck issue*  **-> context vector has to pass all the information through it**
  + **performance can deteriorate**
* *lack of accuracy* **-> does not resolve problem**
* reduce information loss - integrate context vector into each of the decoding vectors
  + improve accuracy
  + bottleneck issue still remains
* Decode by taking entire output from the source sentence as input each time (*assuming GPU supports fast parallel processing*)
* **Weighted sum (w)**  combine attention weights with corresponding values to focus on relevant information -> determine which words to focus on and assign more weight
* enhance performance by updating which words to **give attention** and input accordingly
* decoder can perform calculations required to predict each word in parallel using information from w calculated by encoder

Generative AI (again) - subset of Deep Learning

Drawbacks

* limit to number of tokens that can be processed at once
* built on general-purpose models - lack domain-specific knowledge and understanding for individual industry problems
* performance improves over time and learning and repetition

Algorithm Economy

* evolving state of businesses where algorithms are at the core of value creation
* companies design, build, leverage algorithms
  + differentiate themselves
  + make business processes more efficient
  + create new products/services
  + deliver better customer experiences
  + e.g. algorithm for responding to customer complaints -> speedy and efficient
* value creation being derived from AI algorithms
* growing consensus in business that companies *not* exploring and integrating technological advancements (AI) may find themselves at a competitive disadvantage in the long run
  + tech giants rushing to snap up AI startups
* 4 types of AI companies
* 
* Horizontal AI: provide generalised capabilities across multiple industries or domains
* Vertical AI: designed for specific industries or tasks

Risks of using AI

1. Data Acquisition
   1. difficult to obtain labelled datasets
   2. difficult to recruit experts and collect data
   3. statistical measurement of quality incurs high cost
   4. Strategies to resolve: reinforcement learning, transfer learning
2. Lack of explainability
   1. add more data and use a variety of algorithms introduce bias
   2. not understandable how AI arrives at a result
   3. evolve from simple linear models to non-linear combinations -> black box problem
   4. failure in systems leads to a serious problem
   5. LIME (Local Interpretable Model - Agnostic Explanations)
3. Customer reaction (positive/negative)
   1. anthropomorphism -> attribution of human like qualities to non human entities
   2. uncanny valley -> negative reaction to lifelike robots
   3. perceive robot as less knowledgeable and less empathetic
      1. *fully human like chatbot without disclosure* was perceived to be significantly less likeable than same chatbot incorporating disclosure
   4. users can quickly deduce that the chatbot is not human based on conversational behaviour
   5. ethical concerns + possible solutions
      1. fundamental emotional values and understanding of human experiences affected
      2. core aspects that make human human

Market and Social level Impact of AI

Country vs Company vs Individual

Labour substitution

* create new jobs and displace existing jobs (repetitive and predictable)
* impacts capital intensive and non capital intensive sectors
* need to adapt and remain economically viable
* embrace lifelong learning, T-shape skills
  + combine technical skills with knowledge from other skills, drive innovation in other industries
* leverage flexibility of freelance and remote work to diversify income streams
* reframe meaning of work
  + long term loyalty -> contractual obligations
  + prioritise personal growth and impact of work
* AI patents increase very fast, US dominating

Business level

* robots reduce the need to monitor and supervise workers when robots can reduce human errors in the production process
* Open source software (OSS) depend on contribution of programmers without direct compensation
  + OSS projects create different roles, more employable
  + OSS meritocracy is a credible and precise signal of participants’ productive capacity

Social level conflict due to AI (1x essay question on this)

Intellectual property and AI

* complicated by black box problem of AI
* e.g. Getty Images train millions of pictures without consent from illustrators
* output is vague and controversial
* AI output may be considered copyright infringement if based too closely on existing works, source is not cited, not trustworthy
* important from IP rights holders to be aware of their rights and developments

Hallucination

* **system provides an answer that is factually incorrect, irrelevant or nonsensical, because of limitations in its training data and architecture**
* generate outputs not based on actual data, but errors in model’s programming
* risk of relying on AI systems without proper oversight and testing

Cybercrimes

* ability for threat actors to quickly develop more sophisticated and convincing attacks
  + e.g. model trained on a large dataset of phishing emails to generate new, highly convincing, undetectable phishing emails
* create realistic sounding speech (deepfakes) for social engineering attacks
* automatically generate new security controls or to identify and prioritise vulnerabilities for remediation
* sexual crimes using deepfakes (quote SST example)
* fake yelp review generation
* dual use nature of AI and cyber
  + allow less technically skilled attackers to create more sophisticated cyberattacks
  + can be used for both good and bad purposes
  + underground market exists to assist low skilled cyber criminals

AI regulation (1x essay question on this)

Ensure that AI systems placed on the union market and used are safe and respect existing law on fundamental rights and union values

OECD AI Principles

1. Safety and security
2. Ensuring diversity and inclusiveness
3. Promotion of human rights and fundamental freedoms and human

Essay question: Based on these, how does government take action to protect citizens from AI crime (different from digital crime)

Government regulations

* EU first in the world to pass AI act in May 2023
  + categorise and regulate based on 4 risk levels
  + for high-risk AI systems, government will intervene in business operations
* US introduced AI Bill of Rights - guide design, use and deployment of automated systems
  + safe and effective
  + algorithmic discrimination and protection
  + data privacy
* Japan established AI Systems Council 2024
* South Korea announced Policy Direction for Safe Usage of Personal Data in the Age of Artificial Intelligence in Aug 2023
  + balance innovation with safety and ethical considerations
* Singapore introduced Model AI Governance Framework for Generative AI in May 2024
  + provide organisations with practical guidelines for responsible AI deployment
* AI Crime Regulation
  + need dedicated agencies, ensure equipped to manage and mitigate AI-related crime
  + recruit skilled professionals and adopt technologies that enable automated detection and enforcement
  + provide preventive education on these issues, compensate and support victims to recover from losses
  + Put a pause on AI?