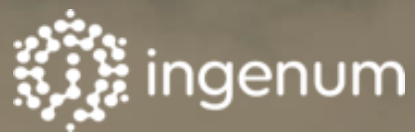


# Applied ML for veterinary epidemiologists

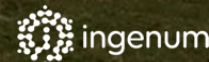
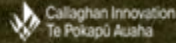
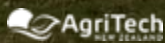
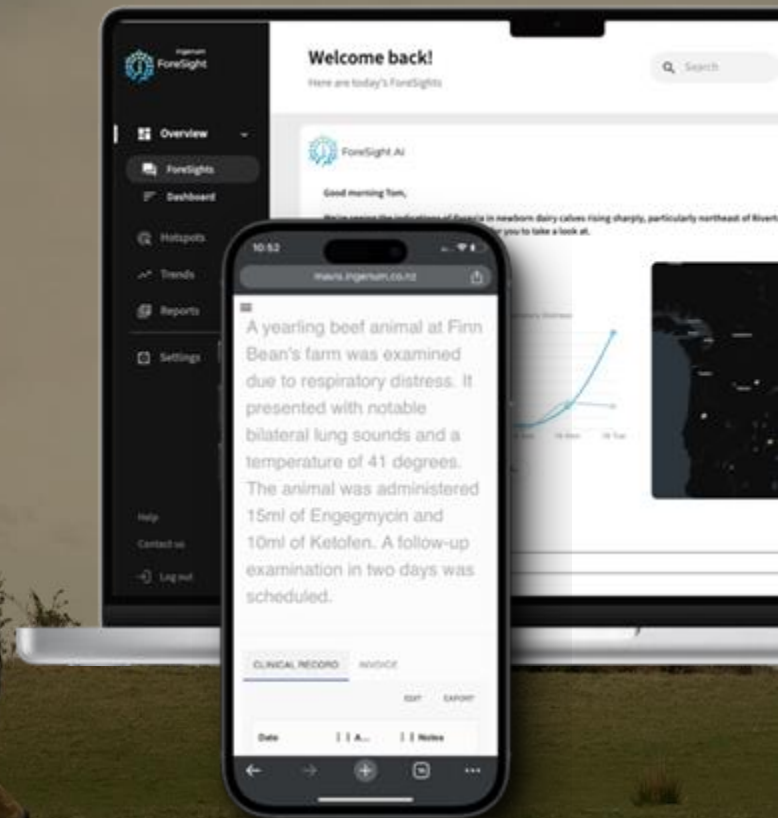
ISVEE pre-conference workshop - Day 4 – Session 1

Dr Tom Brownlie



# Commercial pre-trained for biological pattern detection

Worlds largest animal health dataset



# Ingenium platforms



## Revolutions

1.0

1780  
Mechanisation

Industrial  
production based  
on machines  
powered by water  
and steam

2.0

1870  
Electrification

Mass production  
using assembly  
lines

3.0

1970  
Automation

Automation using  
electronics and  
computers

3.5

1980  
Globalisation

Off shoring of  
production to low-  
cost economies

4.0

TODAY  
Digitalisation

Introduction of  
connected devices,  
data, analytics and  
AI to  
further automate.

5.0

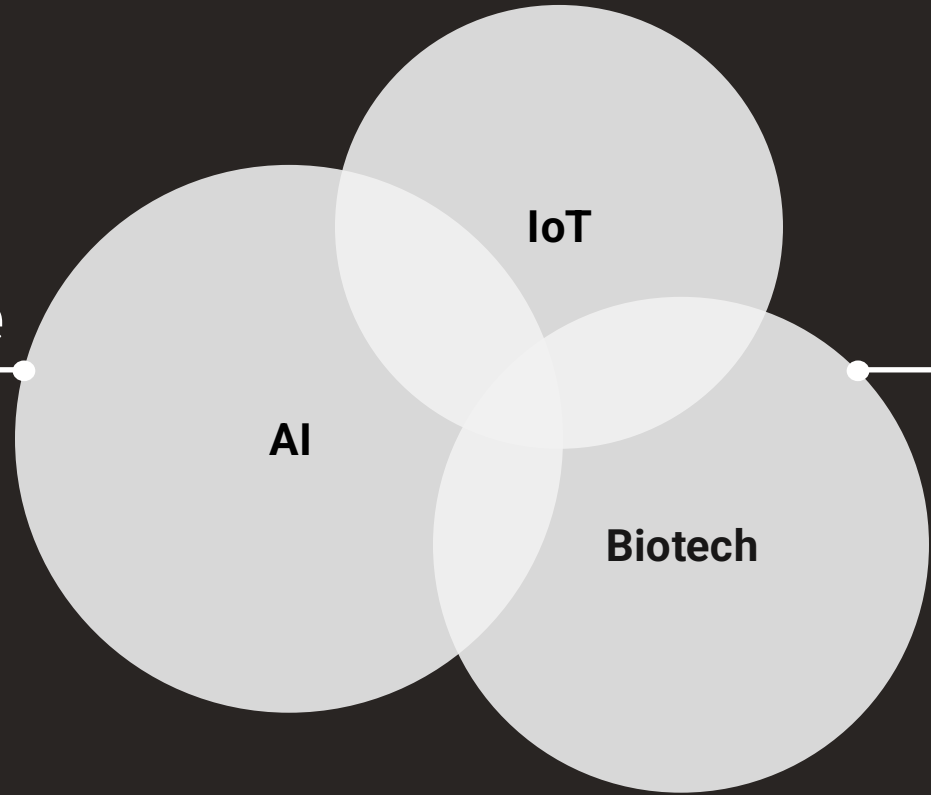
FUTURE  
Human supported  
AI

Humans now a  
supporting role to  
AI powered  
autonomous  
industries

*'Technological revolutions profoundly change the nature of work, occupational structures, and systems of production.'*  
**Barley, S. R.** *What is a Technological Revolution?* Work and Technological Change, Oxford University Press, 2021.

## 4.5 - Technology Supercycle

---



Machine learning is turning things (data) into numbers and **finding patterns** in those numbers.

The computer does this part.

How?

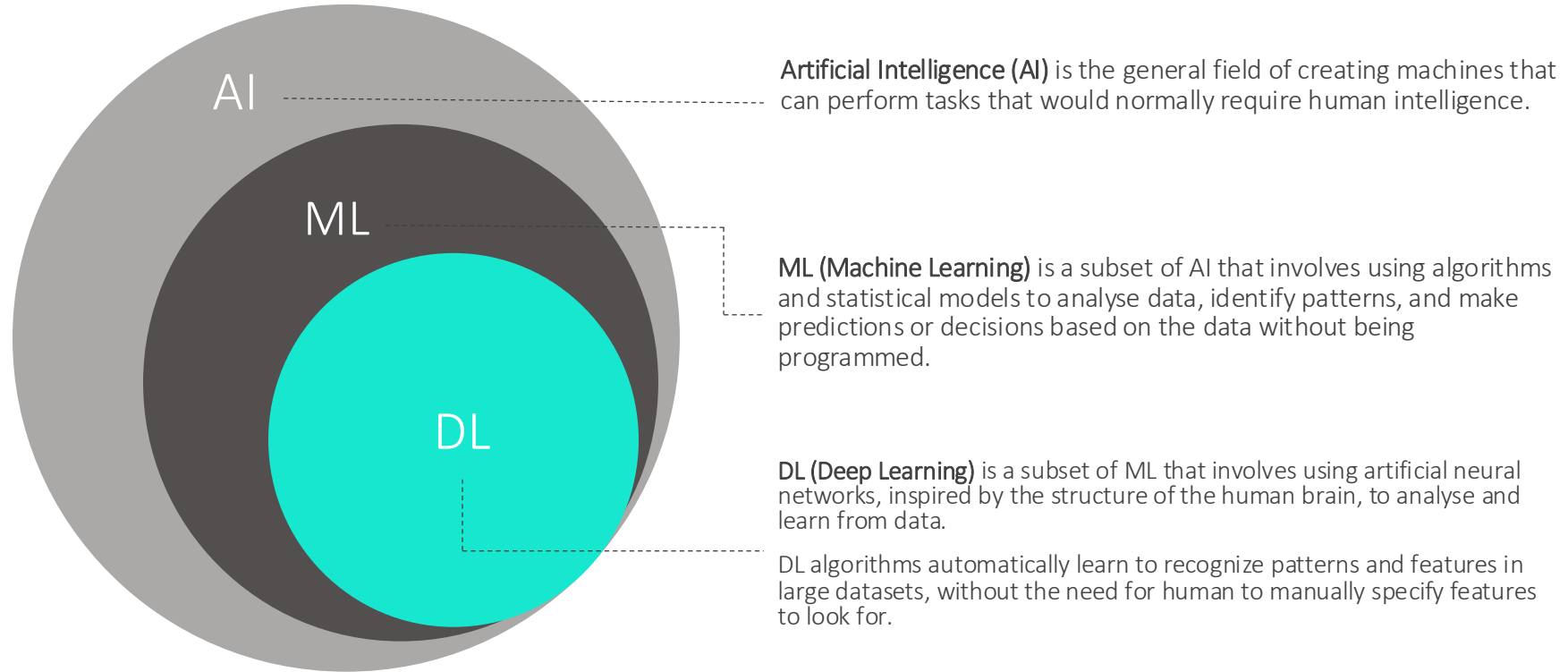
Code & math.

We're going to be writing the code.

# What is deep learning?



# Artificial Intelligence: definitions





# Traditional programming



# ML/DL approach



# Why use machine learning (or deep learning)?



**Good reason:** ~~Why not?~~

**Better reason:** For complex problems  
where you can't imagine all the rules

(maybe not very simple...)

**‘If you can build a **simple rule-based** system that doesn’t require machine learning, do that.’**

Rule #1 Google Machine Learning Handbook

# Deep learning – the good reasons

- Problems with a long list of rules

Where traditional approaches fail, ML/DL may help

- Continually changing environments

DL can adapt (learn) new scenarios

- Discovering insights within large, collections of unstructured data

Imagine crafting rules for all detecting subtle signs of disease in a array of agricultural data?

# Deep learning – the less-good reasons

- When you need **explainability**  
Patterns learned by DP are typically abstract and uninterpretable by humans
- When the **traditional approach is better**
  - Refer to Rule #1
- When **errors are unacceptable**  
Outputs of DL systems are not always predictable
- When you **don't** have much data  
Deep learning models require large amounts of data to produce good results

# ML vs. DL

(common algorithms)

- Random forest
  - Gradient boosted models
  - Naive Bayes
  - Nearest neighbour
  - Support vector machine
- etc.

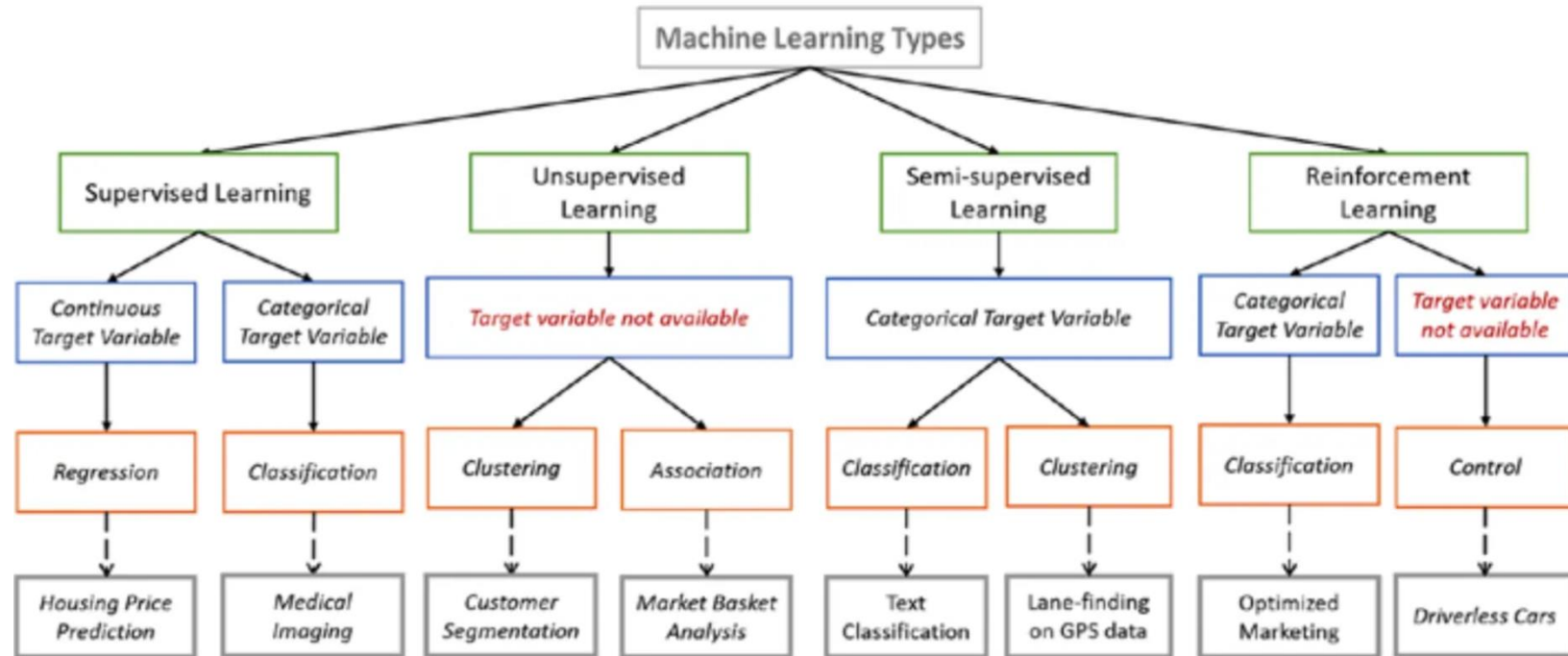
dmlc  
**XGBoost**

- **Fully connected neural network**
  - Convolutional neural network
  - Recurrent neural network
  - Graph neural network
  - Generative Pretrained Transformers
- etc.

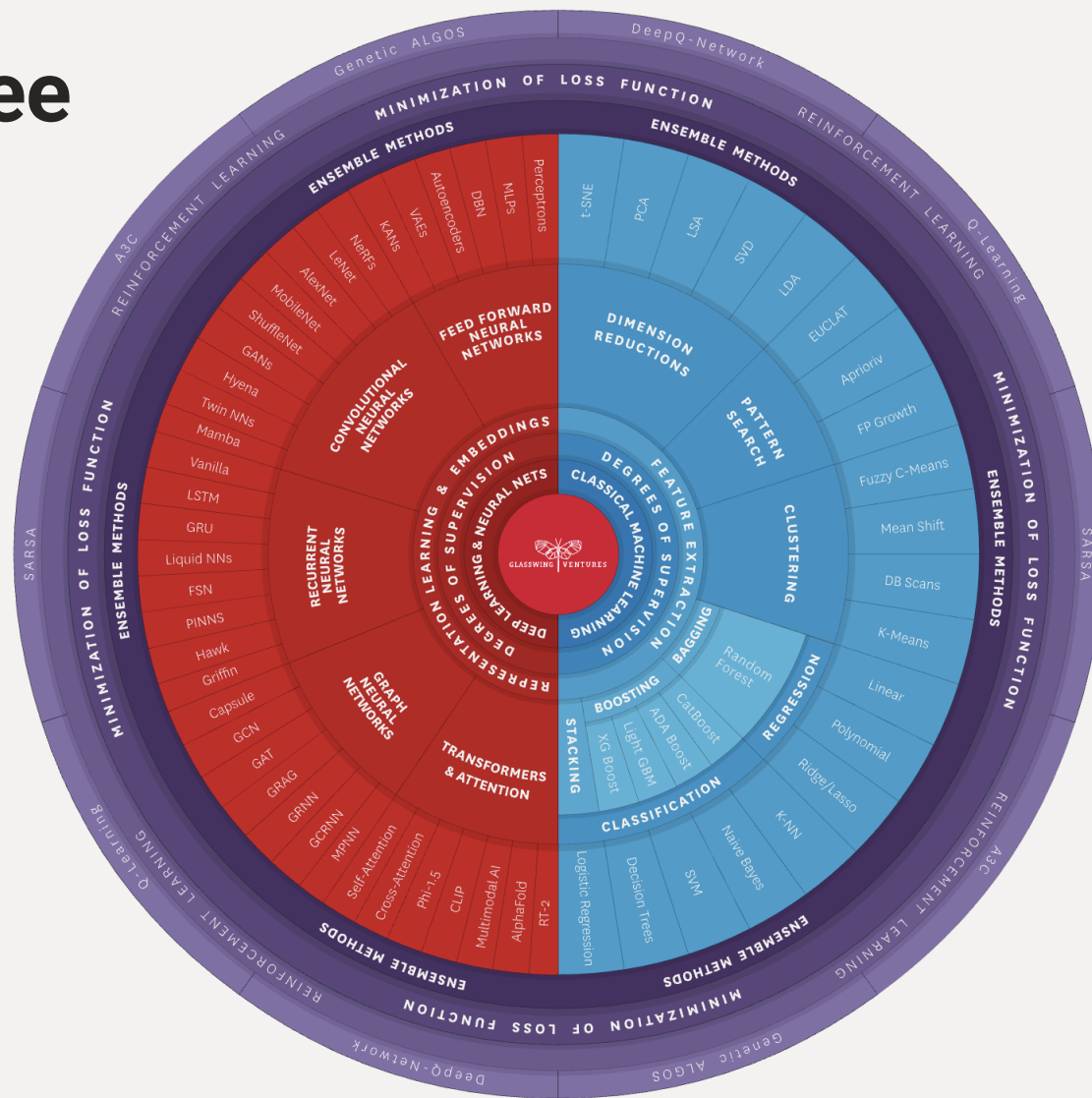
Structured data ← depending how you represent your problem, many algorithms can be used for both → Unstructured data



# Family tree of ML/DL



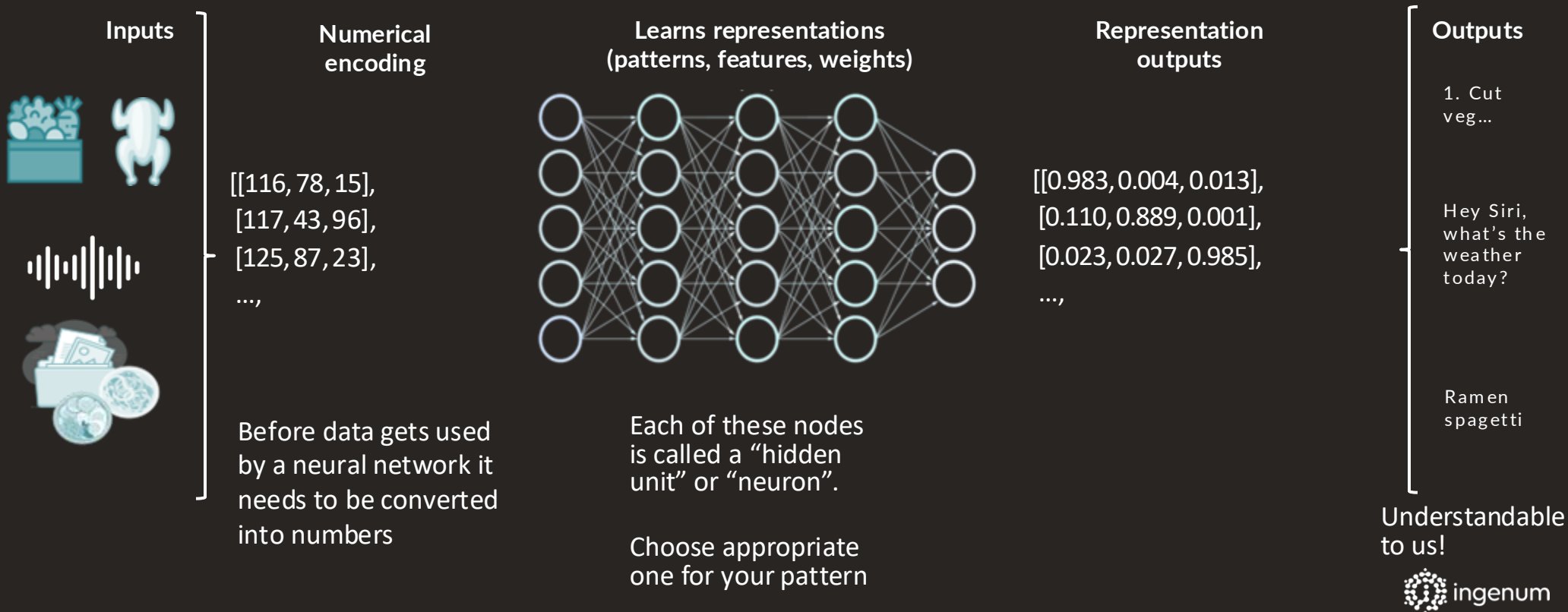
# Family tree of ML/DL



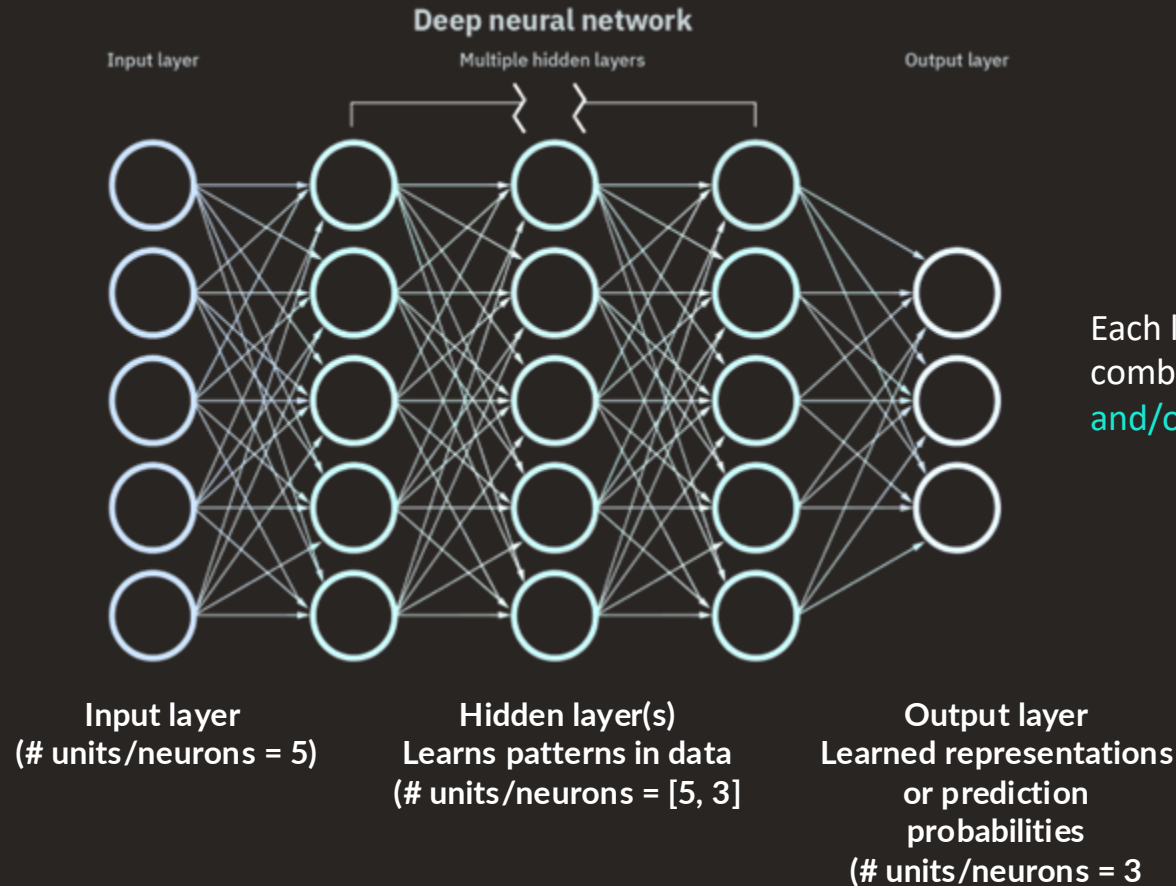
# What are neural networks?

# What are neural networks?

Define real world objects by their connections to other things.

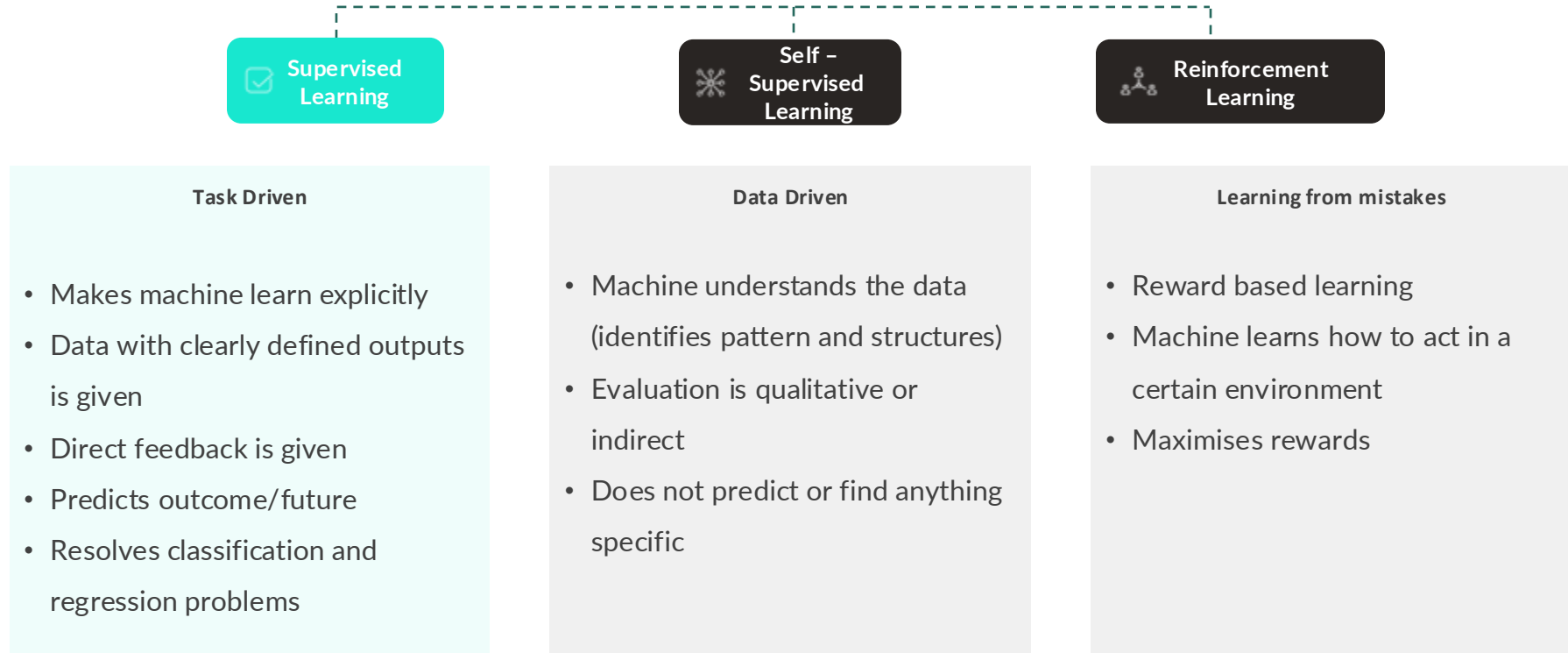


# Anatomy of neural networks



# Deep Learning: model training

Deep learning model training approaches



---

# New toolbox for epidemiologists?



Prediction and  
inference



Clustering  
(segmentation)



Anomaly detection



# What are we going to cover?

# How?

---



GitHub

Repository  
where course  
content is  
stored and  
available



Python

Computer  
language in which  
today's course is  
run



Google Colab

Open-source  
Jupyter notebook  
that contains course  
code  
and Google Gemini  
LLM



Pytorch

Python library  
that supports  
neural networks



Web-based platform for  
version control and  
collaboration

Built in Git (distributed version  
control system)

Hosts, manages and shares  
files in repositories (repo's)

The screenshot displays the GitHub interface for the repository `ingenum-ai / ISVEE_deepLearning_2024`. The repository is private and has 1 star and 0 forks. The main branch is `main`. The repository structure shows folders `images`, `notebooks`, and `supportingMaterial`, along with files `.DS_Store` and `README.md`. The README content includes the title `ISVEE_deepLearning_2024`, a description `ISVEE ML & DL for epidemiology course`, and a draft timetable table.

Time	Format		Brief description
9:00	pptx	Welcome	Introduction to Colab, Markdown and Python
9:20	practical		Introduction exercise in Python
9:50	pptx		Pytorch Fundamentals
10:10	practical		Tensor building exercise

# Python





Python is a general-purpose programming language for data analysis and scientific computing.

Once installed on your computer or server it can be run from command lines/terminal (just like R)

General purpose EDA is done on Jupyter notebooks

More simple code structure than R

Suited to large deep learning models (i.e. Tesla, Microsoft etc)

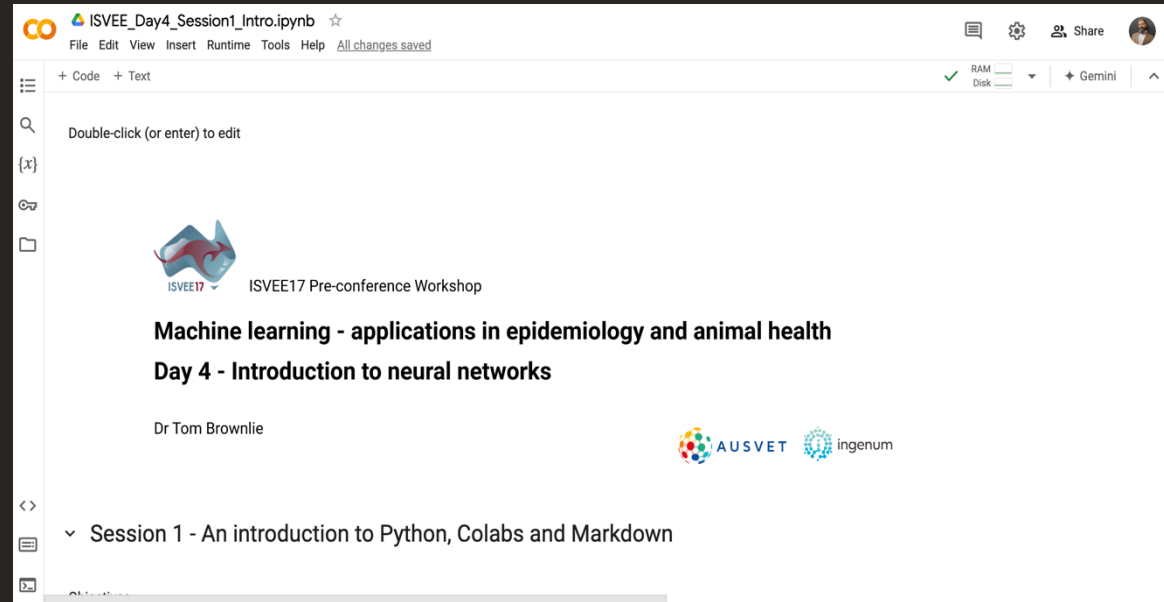
Analysis Tool	Similar Superhero	Super Powers in Common
R 	Batman 	<ul style="list-style-type: none"><li>• Detective Work</li><li>• Intelligence</li><li>• Cunning</li><li>• Usage of Tools</li><li>• More Brain than Muscles</li></ul>
Python 	Superman 	<ul style="list-style-type: none"><li>• Muscle Power</li><li>• Super Strength</li><li>• Elegance</li><li>• Wide Range</li><li>• More Muscles than Brain</li></ul>

# Colab



Google Colab, or "Colaboratory", is a version of a Jupyter notebook that allows you to write and execute Python in your browser, with

- No configuration required
- Access to GPUs free of charge
- Access to Google Gemini (Googles LLM fine tuned to assist with code script)



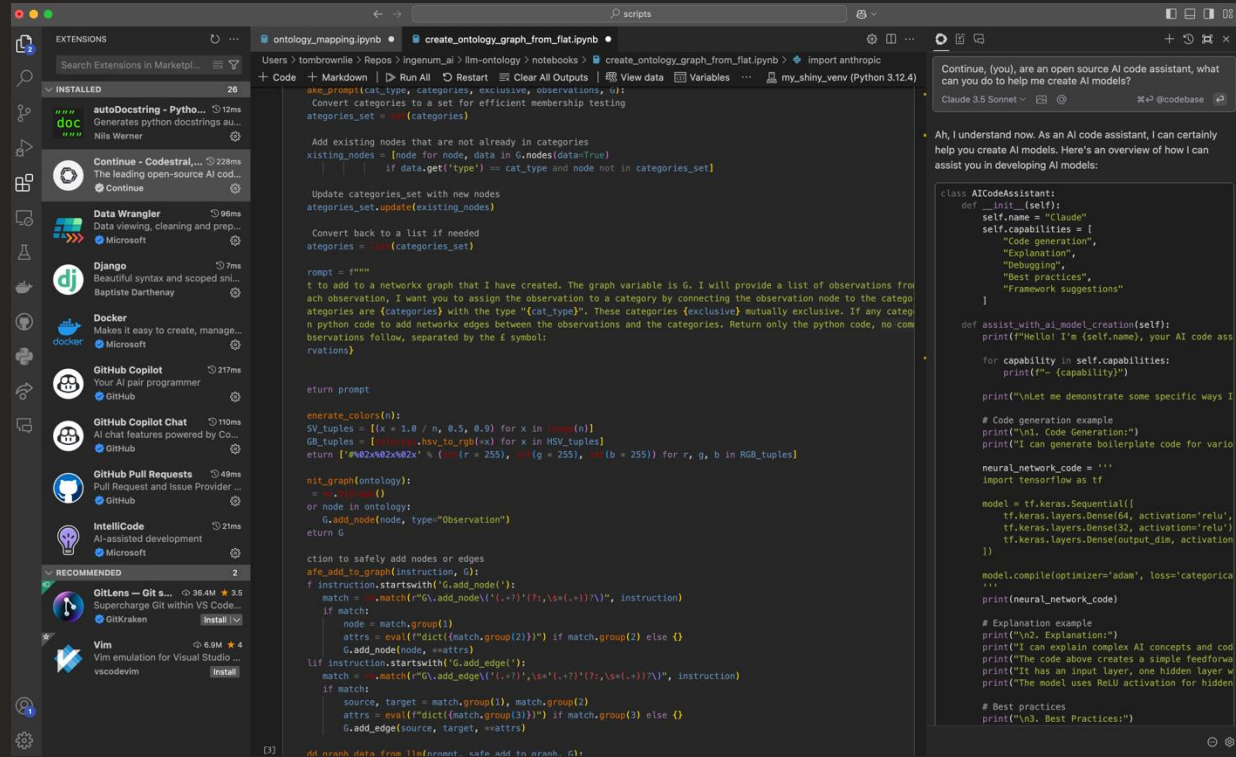
# CO Alternatives

Ingenum use (MS) VS Studio

Other options include PyCharm and Spyder for notebooks

Or simply Notepad++ or Sublime can support python structure

In production environments, python can be 'containerised' (via Dockers) to avoid notebooks







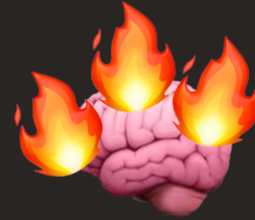
	Python	R
General	general-purpose programming language for data analysis and scientific computing.	Functional programming environment and language for statistical computing and graphics.
Objective	Data Science, Web Development, Embedded Systems	Data Science & Statistical Modeling
IDE	iPython, Pycharm, Jupyter Notebook, Spyder	RStudio, R GUI, R Kward
Data Collection	Supports CSV files, SQL, JSON, and web scraping with BeautifulSoup.	Can also import CSV files with built-in readr library. R's library RCurl provides a simple way to make API requests, similar to Python's requests package.
Data Analysis	Organize dataframes with Pandas for filtering, sorting. Python takes a more streamlined approach for data science projects.	Complex data visualization tools make the exploratory data analysis (EDA) process much more complex than Python.
Essential Packages & Libraries	Numpy, Pandas, matplotlib, scipy, scikit-learn, TensorFlow	caret, stringr, ggplot2, knitr, tidyverse, markdown, shiny, forcats, haven
Database Handling Capacity	Can easily handle large data because there are fewer constraints for memory usage.	R computes everything in memory, so its capabilities are limited by RAM size. A major downfall of R is the inability to handle massive amounts of data.
Data Visualization	Despite the capabilities of data visualization tools like Matplotlib and Seaborn, Python fails to measure up to data visualization features of R.	Developed by and for statisticians, R has complex data visualization features.
Syntax	The "zen of Python" is that there's a proper way to write code.	R doesn't have this set of rules. Also, indexing starts at 1, which can be considered unconventional for general programmers.
Learning Curve	Simple and readable code structure makes it easier for beginners to learn. It also allows for object-oriented programming. It also offers a wide range of data structures that you wouldn't expect from a general-purpose language.	R's functional syntax isn't easy for beginners, but not too challenging for those well-versed in programming. It also offers a few data structures, but fails to handle large amounts of data.



# How to approach today

Each session will follow previous days:

- Short presentation
- Code exercise in notebook
- Questions to extend the exercise (if required)



**How not to  
approach today**

# Resources



Course tutors



Google's in-built  
native LLM



<https://pytorch.org/>



Elon Musk ✓  
@elonmusk

...

Deus ex machine learning

LEARNING ML/DL  
FROM UNIVERSITY

ONLINE COURSES

FROM YOUTUBE

FROM ARTICLES

FROM MEMES



8:07 AM · Nov 18, 2020 · Twitter for iPhone

14.9K Retweets 2.3K Quote Tweets 188.4K Likes

# Lets take a look at a Colab

[https://github.com/ingenum-ai/ISVEE\\_deepLearning\\_2024/](https://github.com/ingenum-ai/ISVEE_deepLearning_2024/)

Open Notebook 1...