Solving the paper to data problem

An introduction to machine learning

How AI can be used with legal contracts in law, financial and professional services

Coverage

- An gentle introduction to AI and its subsets
- Understanding artificial neural networks, and how they actually work
- Use cases in the financial sector
- An introduction to Natural Language Processing, and how it can be used in legal and finance
- Practical examples:
    - data digitisation
    - Basic Q&A on contracts
    - the industrialisation of models
    - A demo of the very latest models, and how they will change how we use AI
- Where to start
- Getting access to the code in the practical examples
- Important take away messages
- Q&A

What is machine learning, and base principles?

Machine Learning is the ability for the machine to learn without being programmed

It relies on three core principles

The data exists	A pattern is thought to exist in the data	It is not possible to pin the pattern down mathematically

Where is machine learning used?

Automotive – eCall for all EU cars from 2017, assisted and fully automated driving, driver behaviour analytics

Automotive Insurance – Black box, Mobile and ODB2 insurance data evaluates drive behaviour to understand elements such as home, content and life insurance

Life insurance – Wearable sensor data feeds back body vital signs, enabling predictive risks

Banking – The banking sector increasingly relies on machine learning for decision making

Healthcare – Genome decoding, cancer treatments, predictive sports injury, actual life expectancy

Professional Services – Converting unstructured data eg legal agreements and financial reports to structured data for analytics

Media and Entertainment – All user inputs are recorded and analysed across the internet, Sky, AmazonTV etc

Weather – Prediction uses some of the largest computers in the world right now

Retail – Amazon claim to know what we want and deliver it before we realise we want it

Gaming – User interactivity analysed and available for sale

Artificial Intelligence timeline

Artificial Intelligence

Techniques to enable computers simulate human behaviour

Machine Learning

Statistical mainly linear models which enable computers to learn without being programmed

Examples include anti-virus and regression based forecasting

Deep learning

A subset of machine learning using artificial neural networks, enabling non linear models to be trained

Examples include complex pattern recognition and generation, and language processing

1950 → 1980 → 2010 →

If you look beyond the marketing hype, this is now a very mature space

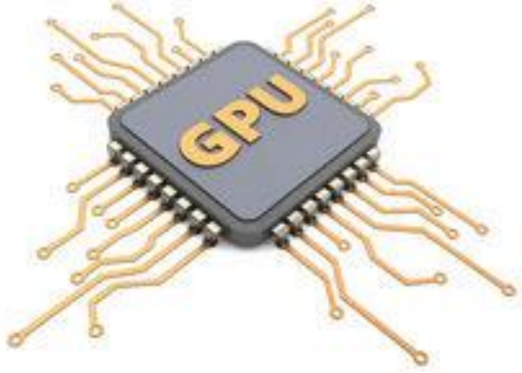
Why now?

Big Data (and lots of it)

Approx 40 zettabytes of data in 2020,

(40,000,000,000,000,000,000,000 bytes)

with 90% of data created in the last two years



Compute power

GPU performance doubling every year, enabling deep learning previously not possible

Brain = 1 Exaflop

Off the shelf NVIDIA DGX-A100 x 6 = 30 Petaflops per rack (33 racks per brain)

In 12 years, a brain on a desktop

IoT (Connected devices)

Connected devices, from mobile devices, wearables, machines, automotive, entertainment and so on

Estimate 31bn by end of 2020, with 127 new devices connected every second

From a human element, what is driving the evolution?

The community is now huge, over 1m users on Kaggle data science community

Anybody can contribute a ground breaking model, and we have competitions that measure the outcomes

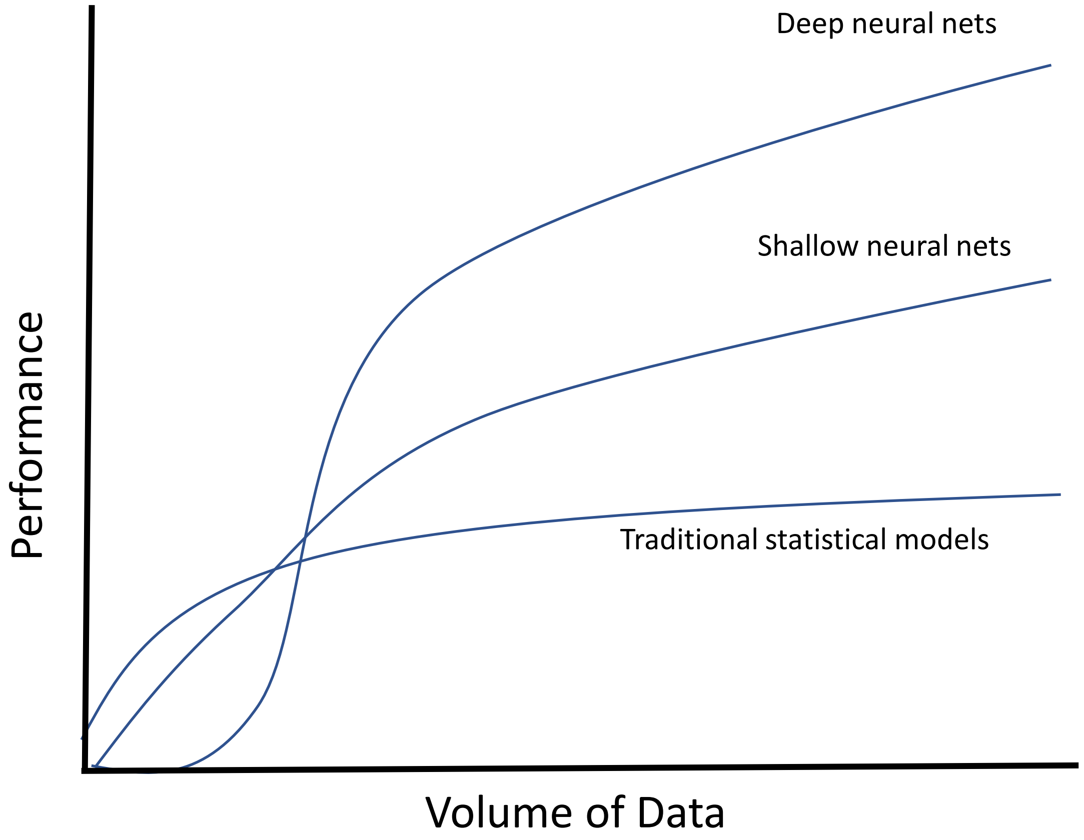
Multi-national companies such as Google and Facebook have open sourced their deep learning frameworks

As soon as a revolutionary advancement is made, it is simplified so practically anybody can use it

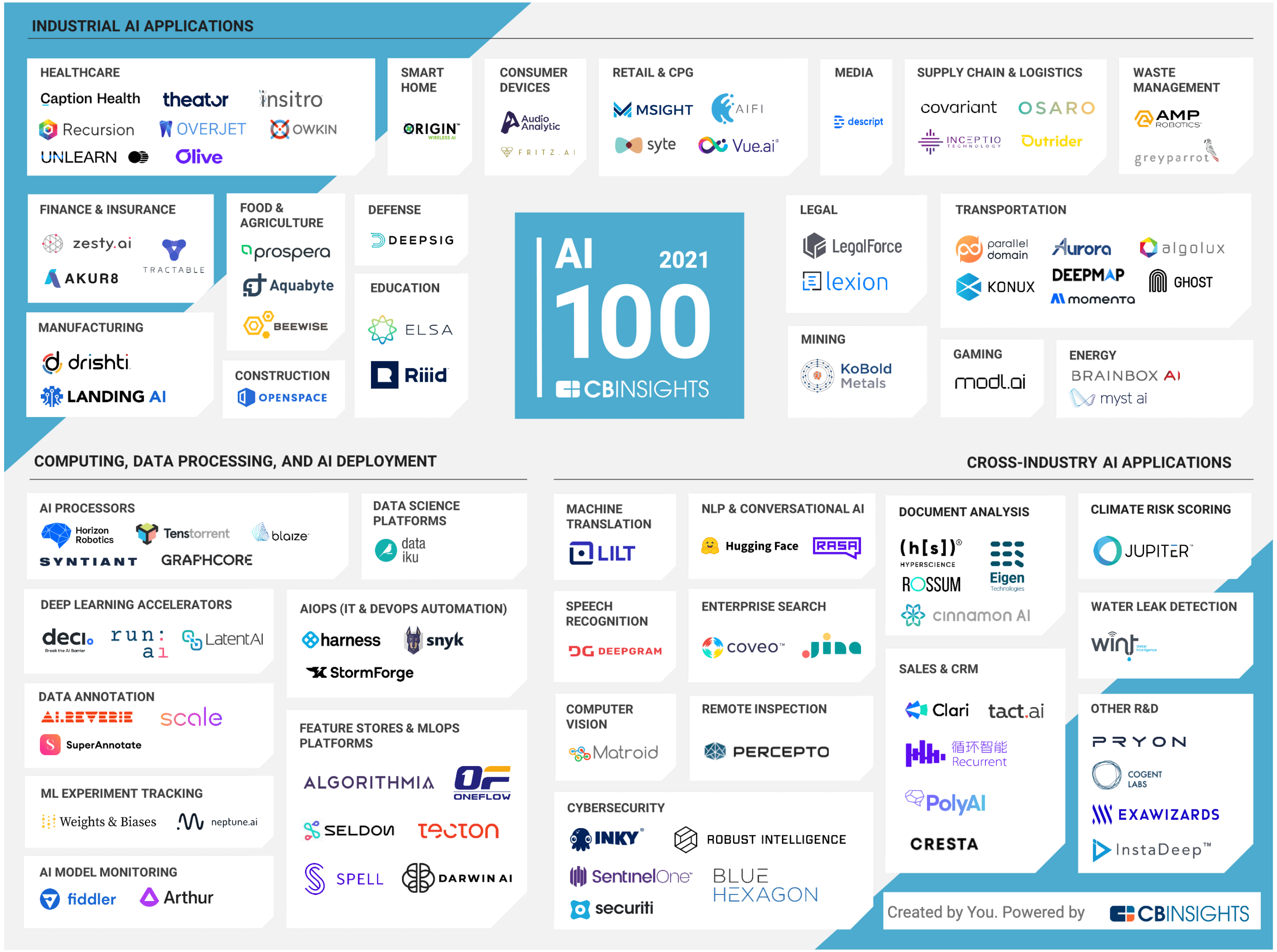
There are sample data sets for just about everything

There are delivery mechanisms for everything over the Internet, including Medium, YouTube, GitHub, Kahn Academy

Statistical models required unique skills, whereas deep learning requires understanding of an ANN and deep learning framework



AI is no longer the domain of academia and 'big tech'



The basics – how do we learn using data

An algorithmic model which learns the relationship between known inputs and outputs using existing data, to enable future predictions

A model where the relationship between known inputs and outputs can be greater understood with more data points and more data, leading to greater prediction accuracy, often far better than humans

Learning the relationships – the basics

What you are used to - A Statistical model

Calculation

Known inputs **+** Known formula **=** Calculated Output

e.g. 2 x 3 = ?

Inputs: 2 and 3
Formula: multiply

A learning model

Training

Predict

Known inputs **+** Formula weights being learned **=** Known Output

Measure gap Between prediction and known output

Update weights in formula

e.g. 2 x Weight = 6

Inputs: 2
Outputs: 6
Formula: multiply

Prediction

A learned model

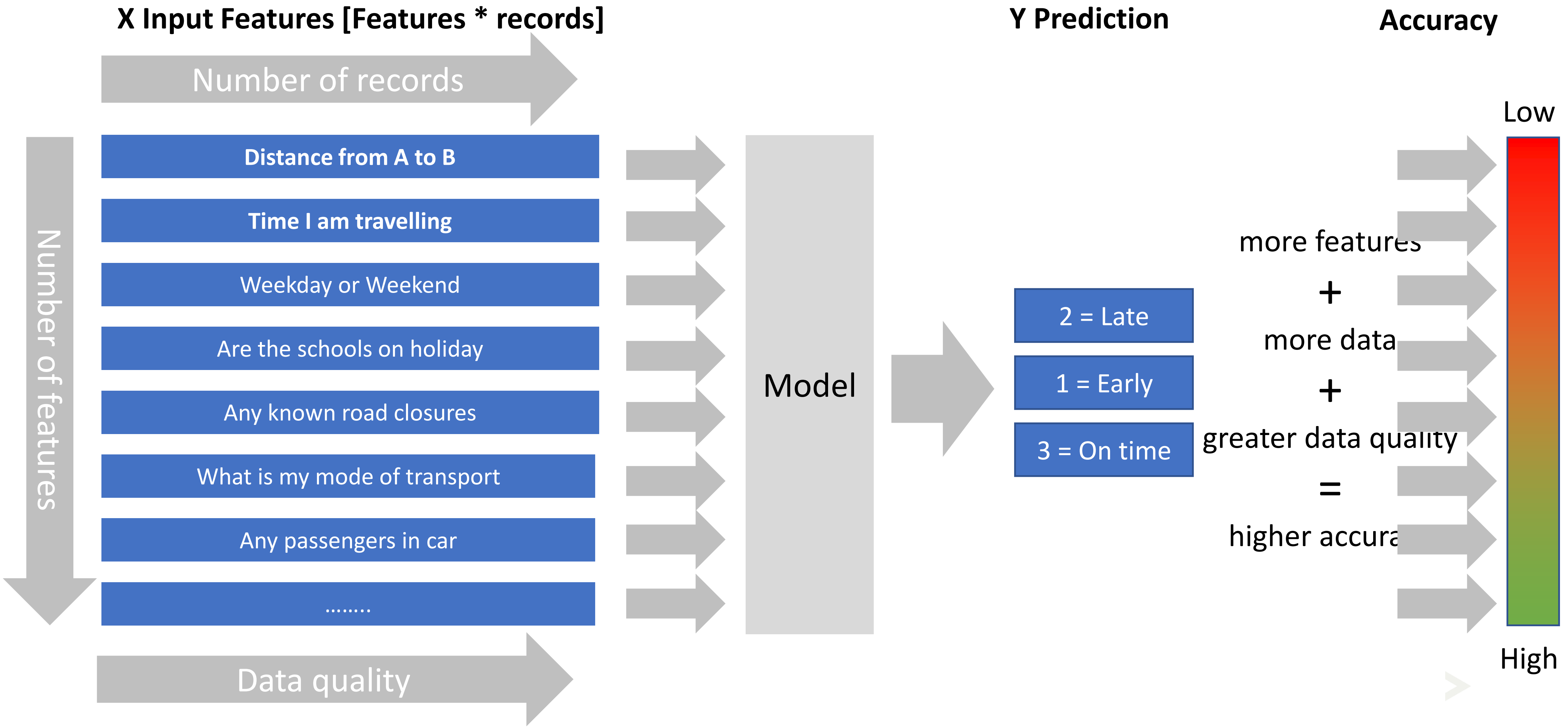
New inputs **+** Formula with learned weights **=** Predicted Output

An algorithmic model which learns the relationship between known inputs and outputs using existing data, to enable future predictions

A model where the relationship between known inputs and outputs can be greater understood with more data points and more data, leading to greater prediction accuracy, often far better than humans

What do we mean by data point, and why they are useful?

Predicting if I can get to work in St Helier in time from Jersey airport using historic traffic data



Feature selection – experts required

Given all the features X is the outcome Y a high probability of a money laundering transaction?

Sample Finance Limited		
Risk Rating		
X Fund Limited		
Country of Nationality	USA	
Country of Operations	UK	
Country Risk Score	**20/100**	Low
Customer Reputation Score	9/20	
Type of Entity	Partnership	
PEP Involvement	Yes	
Type of PEP	Board Director	
Customer Risk Score	**70/100**	High
Type of Industry	Real Estate Development	
Industry Risk Score	**40/100**	Medium
Source of Funds	Qualified Individual Invest	
Source of Funds Risk Score	**50/100**	Medium
Anticipated Account Activity	Multiple per week	
Type of acocunt activity	Loan receipts	
	Payments to suppliers	
	Distributions to investors	
	Loan Payments	
	Receipts from investors	
Account activity risk	**60/10**	Medium
Overall risk score	**48/100**	**Medium**

Sample Finance Limited	
Transaction monitoring	
X Fund Limited	
Activity type	***Data Type***
Date and time request made for payment	*Datetime*
Identity of payment requestor	*Categorical*
Payment recipient type	*Categorical*
Recipient ID for entity	*Numerical*
Payment Destination (Country)	*Categorical*
Payment Coordinates (Payee)	*Numerical*
Destination Account Name (Payee)	*Categorical*
Amount	*Numerical*
Currency	*Categorical*

Type of feature	**Weight**
Change in velocity of payment requests	0.78
New payee with small very frequent payments	0.75
Change in banking coordinates with more frequent payments	0.71
Payee in country different to country of operations	0.69
Investor bank details in country other than residence	0.5
Investor uses shell company in country other than country of residence	0.32
Payment request time of day changes	0.21
Payment Destination (Country)	0.15
Destination Account Name (Payee)	0.15
Date and time request made for payment	0.09
Overall risk rating	0.08
Date and time request made for payment	0.05
Payment recipient type	0.05
Recipient ID for entity	0.05
Currency	0.04
Amount	0.03
Identity of payment requestor	0.02

Feature engineering – Using experts to understand Placement, Layering and Integration

Placement — It is that stage when the “dirty” money is put in the legitimate financial system. The most common way of achieving it is through smurfing, which involves sending small amounts of money to bank accounts that are below anti-money laundering reporting thresholds and later depositing it to the same sender.

Layering — This is the second stage and one of the most complex stages which involves making the money as hard to detect as possible, and further moving it away from the source. The money is purposefully transferred so fast such that the bank cannot detect it.

Integration — The final stage involves putting the “clean” money back into the economy. One of the most common ways is to buy a property in the name of a shell company which shows a legitimate transaction.

Machine Learning Benefits:

- Able to take many features into consideration when determining the probability of Y
- The ‘model’ will keep going until it has reduced its error to the smallest possible amount ie. it will produce the best possible probability given the number of features and number of samples it has.
- The model will show you which features are the most important.
- More features and more samples invariably produces better results
- **Only the best results will be achieved when feature engineering is done by subject matter experts**

Real world example – The power of Artificial Neural Networks over biological

X – Approximately 15bn receptors **Y – Output to other neurons**

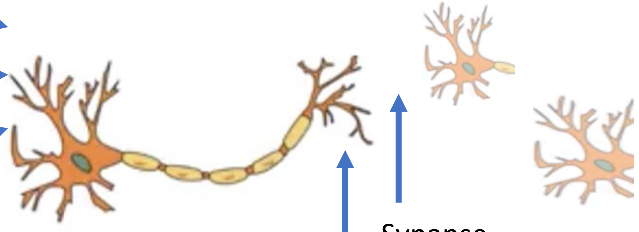
X 100bn neurons

Chemoreceptors

Photoreceptors

Thermoreceptors

Mechanoreceptors



Synapse

Axon terminals

Biological Neuron

Dendrites

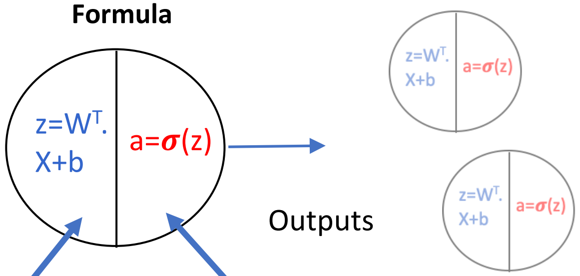
Activation function

X – User defined features **Y – Output to other neurons**

Formula

Text

Images



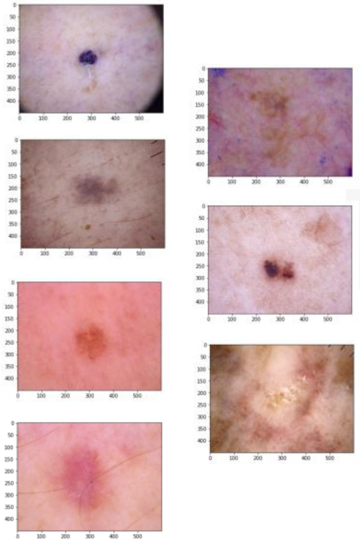
Outputs

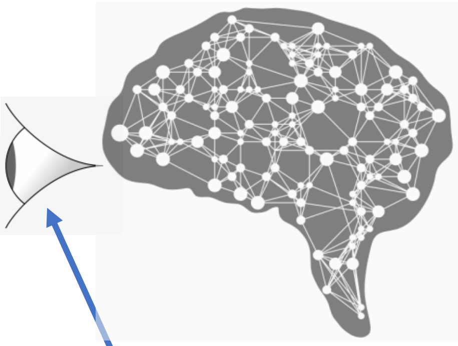
Learn weights

Activation function

Artificial Neuron

X – User defined features





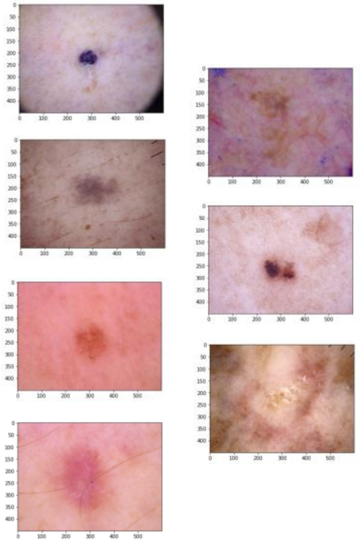
Photoreceptors

Y – Prediction

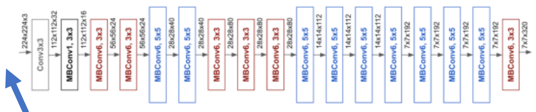
77% accuracy

vascular lesions
Actinic keratoses
basal cell carcinoma
benign keratosis
Dermatofibroma
Melanoma
melanocytic nevi

X – User defined features



Layers of artificial neurons, enabling complex pattern recognition



43 million calculations per image

Photo

Y – Prediction

96.7% accuracy

vascular lesions
Actinic keratoses
basal cell carcinoma
benign keratosis
Dermatofibroma
Melanoma
melanocytic nevi

What do we mean by enabling a model to learn?

Build an untrained brain → Make predictions → Train until smart → Use smart brain to make predictions

Build an untrained brain

Build a model containing a network of artificial neurons with random weights

Make predictions

Make predictions through running the data through the random weights, which generates an output

Train until smart

Update weights in the model through continuing iterations, until the predictions closely match the output

Use smart brain to make predictions

Use the model with learned weights to generate predictions with new data

All data is the same

Data

Numeric	Text	Images	Video	Audio	……	Any sensor data
3.01e-08 2.40e-08 4.44e-08 3.33e-08	Belief, Evidence, Reason, Idea, Claim, Theory, Deduction Think, Test, Connect, Rate, Create, Compare, Generalise Wider, Different, Explicit, Observed, Defined, Deliberate, Thoughtful					
Numeric data can be in many forms including integer, floating point etc	Words are converted to numeric tensors, enabling each word in the vocabulary to maintain a relationship with all other words	Images are converted to any understandable format such as RGB, YCrCb, and then to floating point tensors	Video is the same as images, but time based	Audio is bit rate and sampling over time		Any sensor data to tensors
Features	**Features**	**Features**	**Features**	**Features**		**Features**

Model

Opportunities for Jersey based finance firms and suitable use cases?

The basics

Opportunities and use cases

Serving data and analytics to clients	Jersey's Financial Services customers are mainly other financial institutions, family offices, investment managers and corporates. This client sector has an increasing appetite for analytics. Being able to provide greater value through the provision of data and analytics is valuable.
Automation and increased productivity	A large part of revenue derived from Jersey's Financial Services sector is from selling people's time in a professional services context. Therefore revenue is capped at the amount of the population able to undertake these services. Productivity is capped. In fact the demand for Jersey's finance sector outstrips its supply in this regard and so work is shipped off elsewhere. We know that the key to unlocking standard of living increases comes from raising productivity in any economy. By unlocking productivity in the finance sector we will be able to better serve demand and increase overall standard of living. (1)
Increased competitiveness	Jersey's Financial Services sector can benefit by increasing Accuracy and Efficiency in processes and workflows, thereby becoming more competitive against other jurisdictions.

Paper to Data – Converting latent unstructured data to valuable structured data

- Legal agreements
- Financial statements
- Quarterly Reports

Document and contract review – more efficient in transactions including alternative asset sales and purchases

Fraud detection, error detection and sentiment analysis – providing better risk management

Customer service – being able to automate low hanging customer service needs bringing efficiency and less frustration

Workflow management – AI is better suited to bespoke and narrow use case scenarios than RPA which is typically targeted at high volume processes.

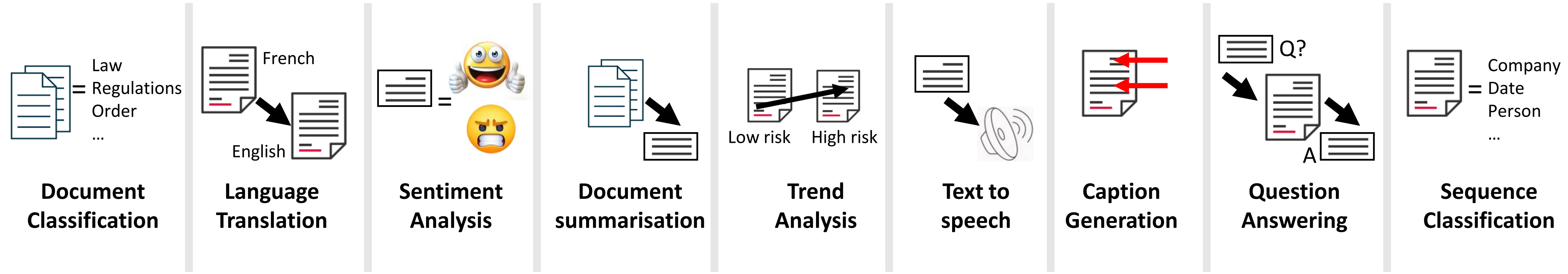
(1) Economic Policy Institute - April 5, 2000. 'The link between productivity growth and living standards.' - https://www.epi.org/publication/webfeatures_snapshots_archive_03222000/

What is Natural Language Processing?

The basics

What is Natural Language Processing?

- It is an area of computer science and AI focusing on the interactions between computers and humans through natural language
  - It involves using specialised techniques to enable the computer to understand raw text data
  - NLP uses a variety of techniques to create structure out of text data
  - Through the understanding of natural language, NLP allows useful interactions such as



Many more uses are appearing every day

Modern NLP Timeline

Neural Language Model 2001

What is it: Probability of the next word appearing based on historic data

Training compute power needed: Tiny, as the calculations only depended on the size of word corpus count, and length of the sequence

Word Embeddings 2013

What is it: Understanding of the relationship between words, such as gender, location. For example: King – Man + Woman = Queen

Training compute power needed: Each word in the vocabulary forming the features has between 100 and 500 dimensions representing each word, but still manageable

Recurrent Neural Networks 2013

What is it: A deep time based artificial neural network to predict or generate sequences of text over time

Training compute power needed: Training possible using GPU’s in modern computers

Attention Models 2015

What is it: An approach to address shortcomings relating to recurrent neural networks long term memory retention, and text alignment

Training compute power needed: Training possible using GPU’s in modern computers

Transformers 2018

What is it: The current standard for natural language processing. Notable models include:
Bidirectional Encoder Representations of Transformers (BERT), used for sequence classification, question answering, named entity recognition
General Purpose Training (GPT), used for sequence predictions
Text to Text Transfer Transformer (T5), used for Q&A, language translation and summarisation

Training compute power needed: If using pre-trained weights, only consumer GPU’s. If training from scratch, think in terms of weeks using the latest GPU technology

Language models are doubling in size every 2.5 months

The latest language models exceed human capabilities

An introduction to Transformers

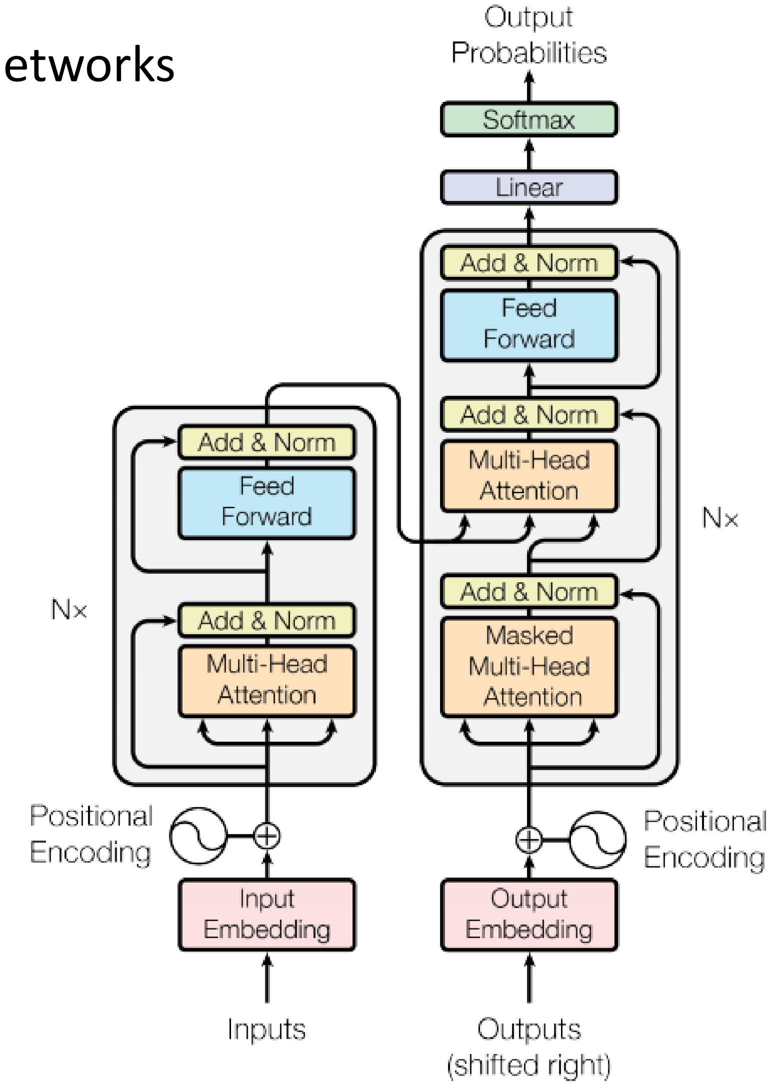
Transformers solve the problem of previous ’forgetful’ recurrent neural networks in NLP

The current state of the art for Natural Language Processing

Uses a concept of ‘Attention’ to learn representations of language models, which can then meet domain specific tasks

Is the foundation for hundreds of pre-trained models

- Generative Pre-trained Transformer (GPT), which is a unidirectional transformer decoder model, used for text generation tasks, including language translation, text summarisation and more
- Bidirectional Encoder Representations for Transformers (BERT), used for token and text classification
- T5 – General purpose Text to Text transformer, and current state of the art



Practical demonstrations

An example of how to digitise your data, password protected, pdf’d or otherwise

- Take away – if you can visually read it, then there is code so anything can read it

Carrying out Q&A on contracts, to enable information to be located from any text almost instantly

- Take away - What once took hours, days and weeks can now be done in seconds with greater accuracy than humans

Industrialisation of your data models, with a practical example of how to extract named entities from documents both now and in the future

- Take away – The data relationships change through time, so it is important to keep your model up to date

A demonstration of the very latest pre-trained universal transformer models

- Take away – Our usage of AI will be outcome based, where we ask for the workings out less and less

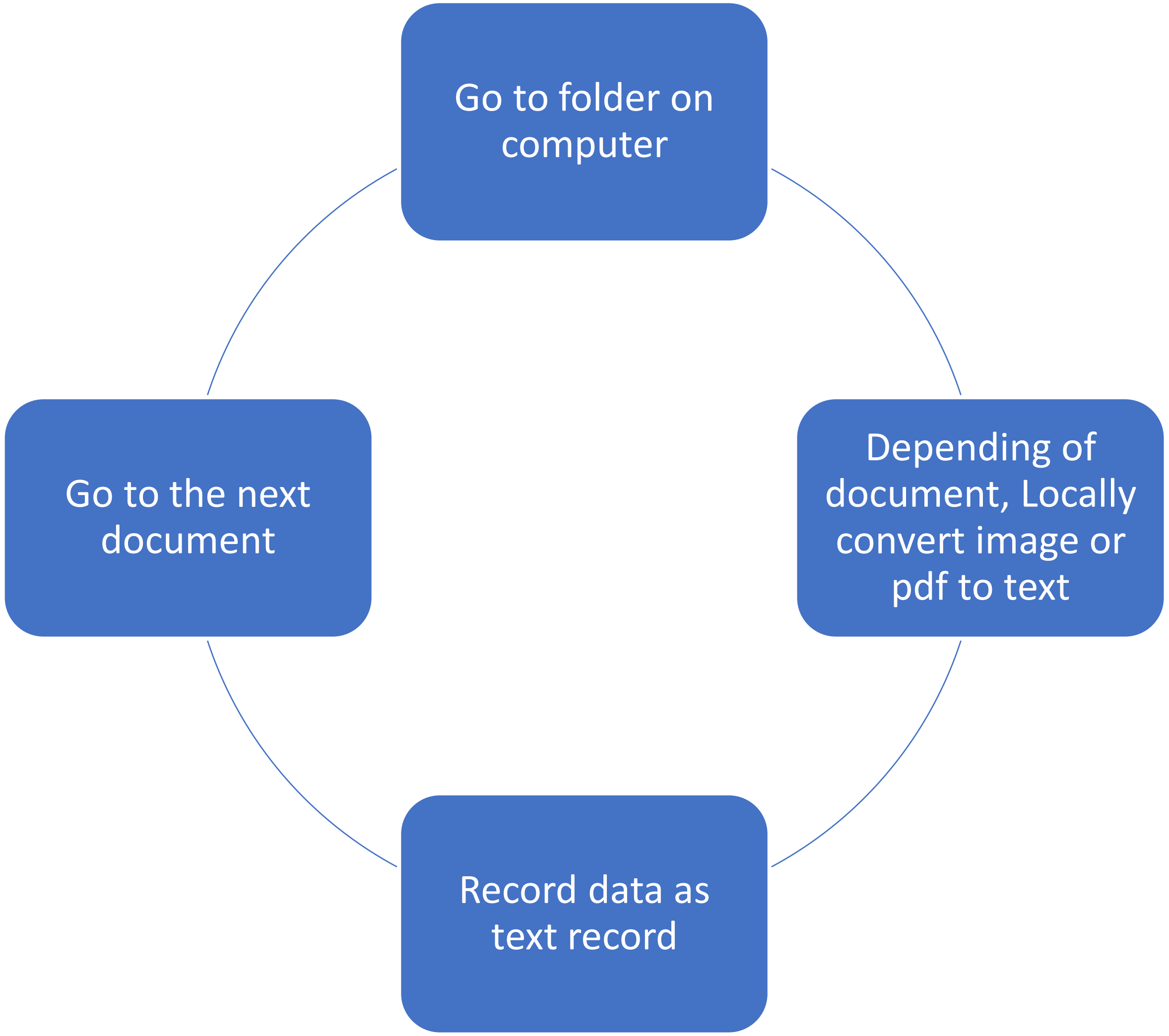
Practical demonstration – Digitise your data

PyTesseract, and its command line variant Tesseract are simple ways to extract text from images locally

Many tools can do the same, depending on document and data type

All types of tools exist to quickly and simply digitise data. If the data source is standard, then there are several options available

NOTE: If you have documents you can physically read, then document protection in word, pdf or otherwise is pointless. OCR can read it



Useful libraries for extracting data

Many Python libraries available to easily digitise your data

PyTesseract – Extract text from images

pdfplumber – Extract text and tabular data from PDF’s

PyPDF2 – Extract text from PDF, split and join docs, and decrypt/encrypt

BeautifulSoup – Web scrape text from the web

spaCy – pre-trained NLP pipelines

nltk – word tokenizer

Practical demonstration – Q&A on legal text using BERT

What exactly is a BERT Model

BERT is a pre-trained language model using the Transformer Encoder to carry out language tasks

BERT is pre-trained using a masked language model (MLM) and next sentence prediction (NSP) concurrently

BERT is pre trained on the whole of Wikipedia 2.5Million words) and Book Corpus (800k words)

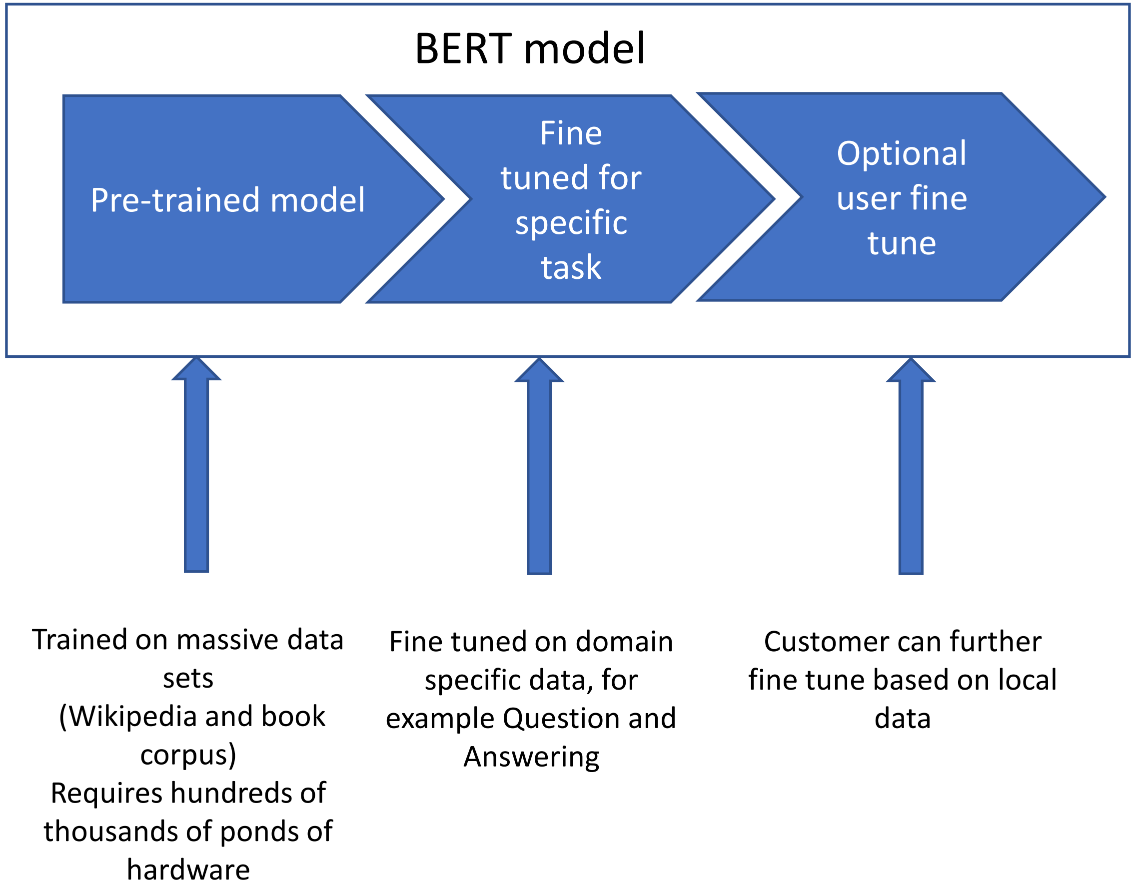
It is further fine tuned to specific tasks using large solution specific data

Fine tuned models exist for:

- Token Classification (Named Entity Recognition and Q&A)
- Text Classification, including sentiment analysis

NOTE: It has limitations, as any record it processes cannot be more than 512 token

NOTE2: Many pre-trained BERT like models have been open sourced by the community based on industry need, however you can fine tune your own based on your own data.



Practical demonstration – Q&A on legal text using BERT

Aims: Extract data and location from legal text, for example Contract sign, Parties, Term etc

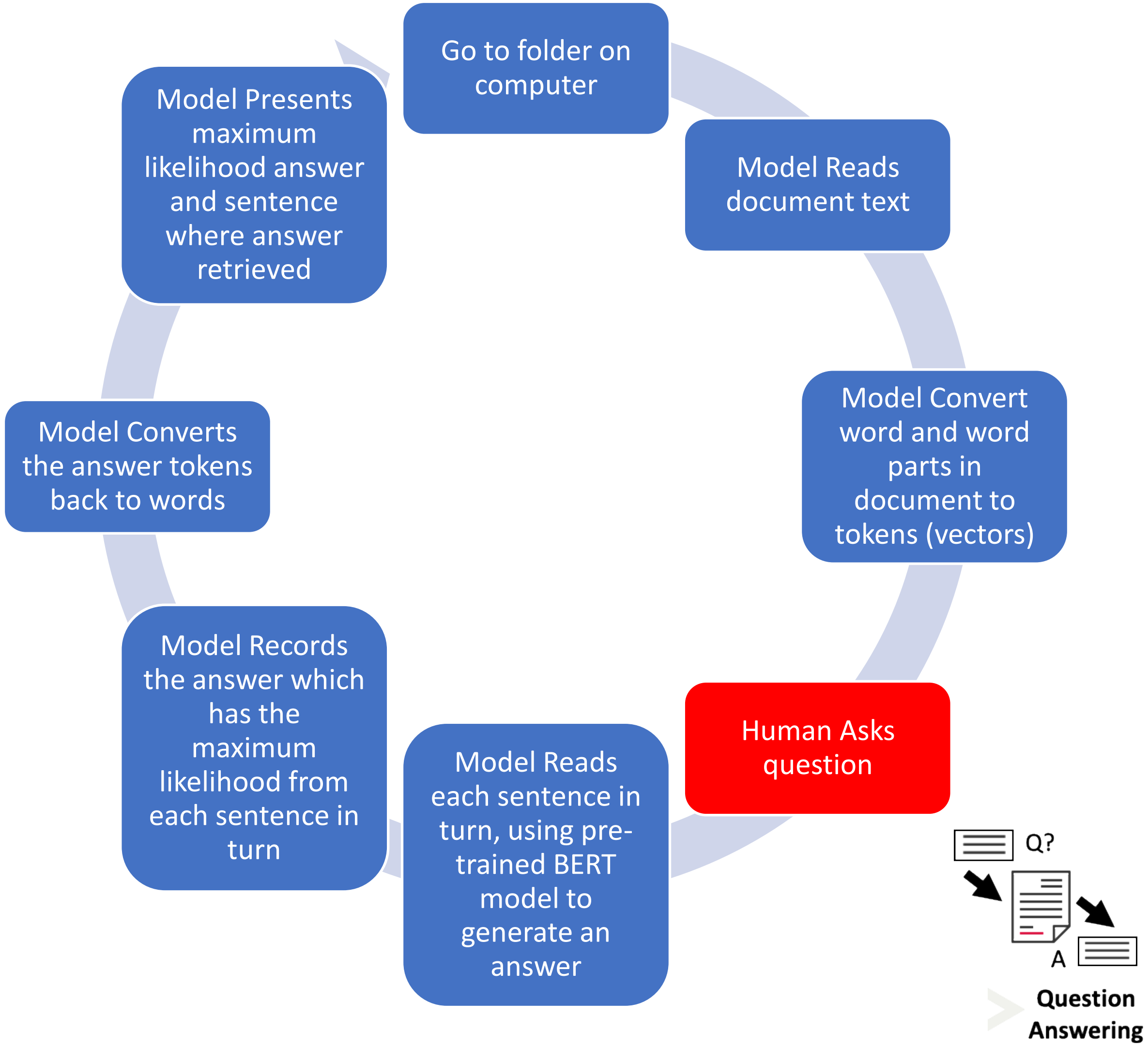
This model is fine tuned for Q&A using a Stanford Q&A dataset (SQuAD) containing 100,000 questions, where the answer is in the context text

The pre-trained weights take up 1.5Gb, and contain approx 340m parameters

Is approximately 80% accurate out of the box

What it is actually predicting is the probability of the start and end of text within the context data

Context	Question	Predicted Answer
We have found that in the summer, it is generally warm, whereas winter is cold.	When is it warm?	**Start token:** 20 **End token:** 32



Fine tuning BERT notes

Expected data in CSV with the below headers:

context – The overall text where the answer exists

answer – The answer text, which is in the **context**

question – Question being asked

answer_start – integer where the text starts

Fine tuning learns **answer_start** and model generated **answer_end tokens** through training

Once the new model is saved, it can be used in future predictions

context	question	answer	answer_start
We have found that in the summer, it is generally warm, whereas winter is cold.	When is it warm?	in the summer	20

Case Study - NLP for data extraction.

How AI can be used to extract key information from your existing documents

Example – How AI can be used to extract key information from your existing documents

So what's the problem?

Motivation:

Companies, law firms, regulators and government agencies need to analyse and monitor contracts for a wide range of tasks.

For example law firms need to review existing contracts when there are amendments, changes in law or when clients need them to review existing contracts.

Companies need to keep track of dates, terms, amounts and parties across many contracts.

The majority of in-house legal departments at mid-sized companies spend 50% of their time reviewing contracts. [(1)] During transactions and due diligence, this is outsourced to external law firms. Billing rates for lawyers at large law firms are typically around $500-$900 per hour in the US.

As a result, many transactions cost companies hundreds of thousands of dollars just so that lawyers can verify that there are no problematic obligations or requirements included in the contracts. Contract review can be a source of drudgery and, in comparison to other legal tasks, is widely considered to be especially boring.

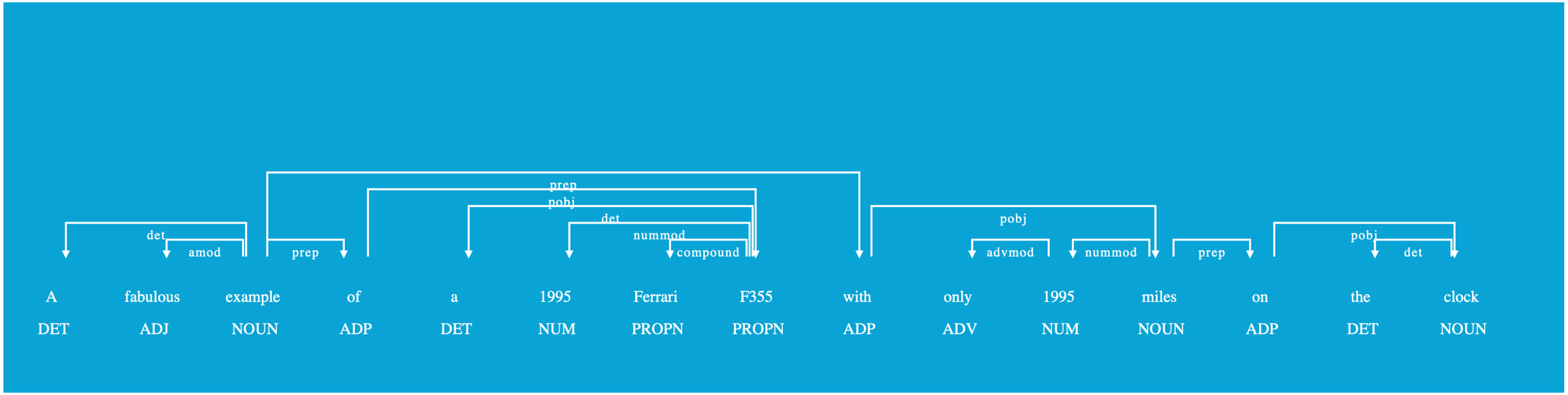
Therefore, on the basis that contract element extraction can be automated, this will save time and money is review, minimise risk of contract default and provide useful data to be used in analytics.

(1) Gartner,2017 - https://web.archive.org/web/20170920135124/https://www.cebglobal.com/compliance-legal/smb-legal/contract-management-midsized.html

Example – How AI can be used to extract key information from your existing documents

Pre-trained Named Entity Recognition software libraries such as spaCy and Transformers are already excellent at understanding context

1. It's confirmed that Apple **ORG** is looking at buying U.K. **GPE** startup for $1 billion **MONEY**
2. Apples are available from the local Co-Op **ORG** in Jersey **GPE** for only 30 pence **MONEY**
3. A fabulous example of a 1995 **DATE** Ferrari **ORG** F355 **PRODUCT** with only 1995 miles **QUANTITY** on the clock



Example – How AI can be used to extract key information from your existing documents

But this is a domain specific problem.

SERVICES AGREEMENT 1

This agreement is dated 20 April 2021 2

Parties

1 3 **Whiterock Investment Management Limited,** incorporated and registered in England and Wales with company number 456123 whose registered office is at Queens Towers, 50 Pall Mall, London ("**Whiterock**").

2 3 **NATIVE LIMITED** incorporated and registered in Jersey with company number 130087 whose registered office is at Le Capellain House, Jersey, JE3 5UT ("**Provider**").

Agreed terms

1 Interpretation

1.1 The following definitions and rules of interpretation apply in this agreement (unless the context requires otherwise).

Board: the board of directors of Whiterock (including any committee of the board duly appointed by it).

Commencement Date: 20 April 2021 2

Generic named entity recognizers (NERs) typically recognize persons, organizations, locations, dates, amounts, etc., which are unfortunately not directly applicable.

For example, a generic NER may recognize dates, but without distinguishing between start, effective, termination and other dates.

Similarly, for data extraction, we only want legal entities which are party to the contract to be identified, not defined terms (such as "Provider" or " Whiterock") or other referenced legal entities.

Example – How AI can be used to extract key information from your existing documents

How is research and collaboration solving this domain specific problem?

2017

- Athens University of Economics and Business and Cognitiv+, a London company specialising in Artificial Intelligence solutions for document review and analysis and regulatory compliance, releases a research paper named ‘Extracting Contract Elements’ together with a new benchmark dataset. This attempts to utilise standard machine learning and NLP techniques to automate contract element extraction. (1)

2020

- Following the ground-breaking developments around Transformers in 2018 resulting in the publication of BERT in 2019, Athens University, the Greek National Centre for Scientific Research and the University of Sheffield publish a research paper on Pre-Training and Fine-Tuning Transformer models specifically for the legal domain including contracts, statute and case precedents. (2)

22 February 2021

- Athens University, the Greek National Centre for Scientific Research publish ‘Neural Contract Element Extraction Revisited‘ following the advancements in Transformers over the past two years. (3)

10 March 2021

- The Contract Understanding Atticus Dataset (CUAD) was introduced , a new dataset for legal contract review. CUAD was created with dozens of legal experts from The Atticus Project and consists of over 13,000 annotations. (4)

Expected in Summer 2021

- Large expansion of the CUAD database from Atticus with the objective to enhance the performance of the fine-tuning of the state-of-the-art Transformer models.

(1) Ilias Chalkidis, Ion Androutsopoulos, and Achilleas Michos. 2017. Extracting Contract Elements. In Proceedings of International Conference on Artificial Intelligence and Law, London, UK, June 12–15, 2017 (ICAIL’17), 10 pages. DOI: http://dx.doi.org/10.1145/3086512.3086515

(2) arXiv:2010.02559 [cs.CL]

(3) arXiv:2101.04355 [cs.CL]

(4) arXiv:2103.06268v1 [cs.CL]

Example – How AI can be used to extract key information from your existing documents

How do we go about creating the ‘Feature Labels’ to fine-tune an existing state-of-the art NLP model?

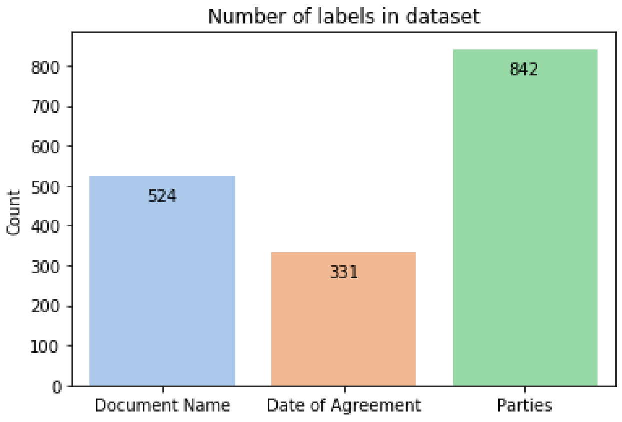
STRATEGIC ALLIANCE AGREEMENT DOC_NAME This Strategic Alliance Agreement DOC_NAME (this “Agreement”) is made and entered into as of May 17, 2016 AGMT_DATE , (the “Effective Date”) by and between Giggles N’ Hugs, Inc. PARTY , a Nevada corporation (the “Company”), and Kiddo, Inc. PARTY , a Florida corporation, Inc., (“Consultant”). WHEREAS, Consultant is retained by the Company as an Independent Contractor to introduce investors, celebrity spokespersons, press and media relationships, raise public awareness of the company and its public securities, and for other services related to Consultant’s expertise; and WHEREAS, the Company and Consultant have agreed upon, and wish to memorialize their agreement concerning the status and responsibilities of the parties. NOW, THEREFORE, the parties agree as follows: 1. Services (a) General. Consultant shall use all best efforts to provide services including the following: ● Capital ○ Introduction to key investors. ○ Introduction to strategic partners ○ Introduction to

Source – Docanno annotation library

Example Feature Labels

```
0 B-AGMT_DATE
1 B-DOC_NAME
2 B-PARTY
3 I-AGMT_DATE
4 I-DOC_NAME
5 I-PARTY
6 O
```

```
The total number of labels in the dataset is: 1697
```



id	ner_tags	split_tokens
5594	[6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 1, 4, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 0, 3, 3, 3, 3, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ...]	[1, EXHIBIT, 10.26, CONFIDENTIAL, MATERIALS, OMITTED, AND, FILED, SEPARATELY, WITH, THE, SECURITIES, AND, EXCHANGE, COMMISSION, ., ASTERISKS, DENOTE, OMISSIONS, ., SPONSORSHIP, AGREEMENT, This, agreement, (, ", Agreement, ",), is, entered, into, as, of, the, 23rd, day, of, September, 1998, (, ", Effective, Date, ",), ,, by, and, between, Excite, ,, Inc., ,, a, Delaware, corporation, ,, located, at, 555, Broadway, ,, Redwood, City, ,, California, 94063, (, ", Excite, ",), ,, and, Vitamin, Shoppe, Industries, Inc., ,, a, New, York, corporation, ,, located, at, 4700, Westside, Avenue, ,, North, Bergen, ,, New, Jersey, 07047, (, ", Client, ...]
5831	[1, 4, 4, 4, 4, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 0, 3, 3, 3, 3, 6, 6, 6, 6, 6, 6, 6, 6, 2, 5, 5, 5, 6, 6, 6, 6, 6, 6, 2, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 2, 5, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ...]	[Consulting, and, Product, Development, Agreement, ARTICLE, 1, PREAMBLE, This, Consulting, and, Licensing, Agreement, (, ", Agreement, ",), is, entered, into, this, 1st, day, of, September, 2016, (, “, Effective, Date, ",), by, and, between, Emerald, Health, Sciences, Inc., (, “, EHS, ",), ,, Emerald, Health, Nutraceuticals, Inc., (, “, EHN, ",), ,, and, Michael, T., Murray, ,, N.D., (, “, Dr., Murray, ",), ., This, Agreement, sets, forth, a, description, of, those, responsibilities, of, EHS, ,, EHN, ,, and, Dr., Murray, ,, of, certain, rights, granted, to, EHS, and, EHN, ,, and, of, certain, other, ...]
5660	[6, 6, 6, 6, 6, 6, 0, 3, 3, 3, 2, 5, 5, 5, 6, 2, 5, 1, 4, 6, 6, 6, 6, 6, 6, 6, 6, 6, 0, 3, 3, 3, 6, 6, 6, 6, 6, 2, 5, 5, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 2, 5, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, ...]	[Exhibit, 10.2, STRICTLY, PRIVATE, AND, CONFIDENTIAL, 1, April, ,, 2020, THERAVANCE, BIOPHARMA, UK, LIMITED, and, BRETT, HAUMANN, SERVICE, AGREEMENT, THIS, AGREEMENT, is, entered, into, between, the, parties, on, 1, April, ,, 2020, ., PARTIES, (, 1,), Theravance, Biopharma, UK, Limited, is, a, company, registered, in, the, United, Kingdom, and, whose, registered, office, is, at, 12, New, Fetter, Lane, ,, London, ,, United, Kingdom, ,, EC4A, 1JP, (, the, “, Employer");, and, (, 2,), Brett, Haumann, of, [, address, removed,], (, the, ", Executive, ",), ., AGREED, TERMS, 1, ., Definitions, 1.1, The, following, terms, shall, have, ...]

Example – How AI can be used to extract key information from your existing documents

Now that we have enough documents and enough ‘Feature Labels’ we can fine-tune an existing model.

1. Pre-process the documents (the dataset).

Convert to text from PDF or images.

‘Clean’ the text so it is easy to label each word in each document.

2. Annotate the dataset

Tag each important word in each document with a chosen feature label schema, essentially reflecting the data you want to extract

3. Use your new dataset to ‘fine tune’ a pre-trained state-of-the-art NLP model

This makes a very smart model just a little smarter by adding your own data

Here you also get to choose what you are trying to optimise, this is important.

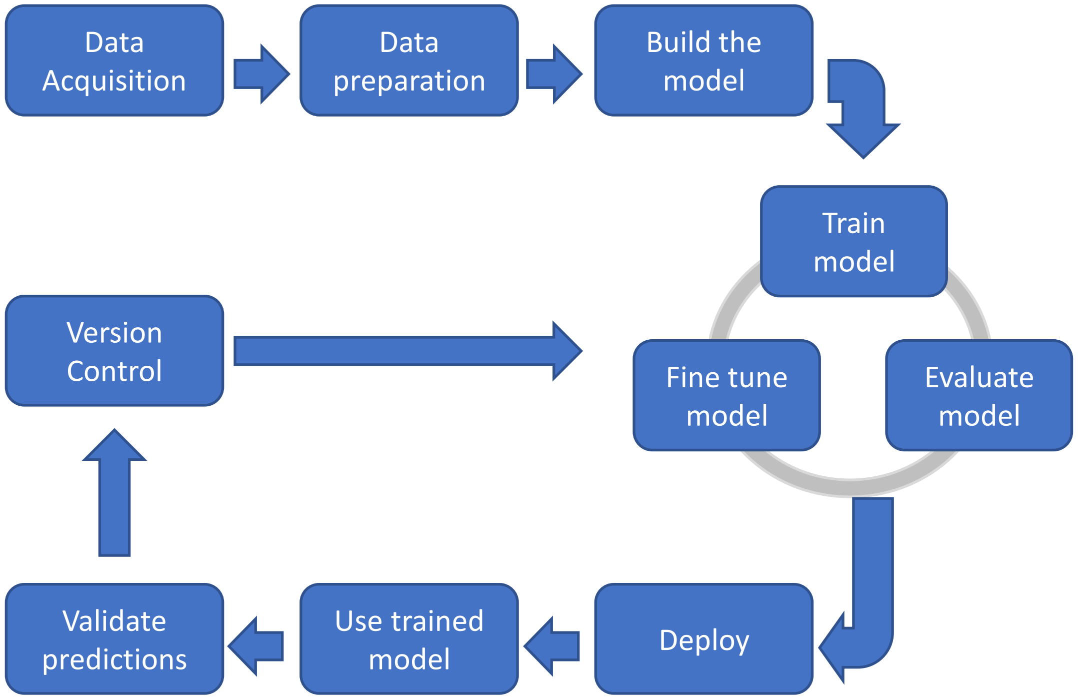
Use this new fine-tuned model to extract data

Keep checking the performance and feedback any corrections to the training dataset to make the model smarter and adapt to change

Practical demonstration - industrialisation of your data models using an named entity recognition example

Your data will change through time, aligned to your industry

Industrialisation of your data strategy through time is critical to keep AI as accurate as possible



Current state of the art – T5

A summariser, Question Answering model and Language Translator in one model

Practical demonstration – Q&A and document summary on legal text using T5

Developed by Google, T5 is the very latest experimental Transformer model to carry out document summarisation, question answering and language translation in one massive pre-trained model

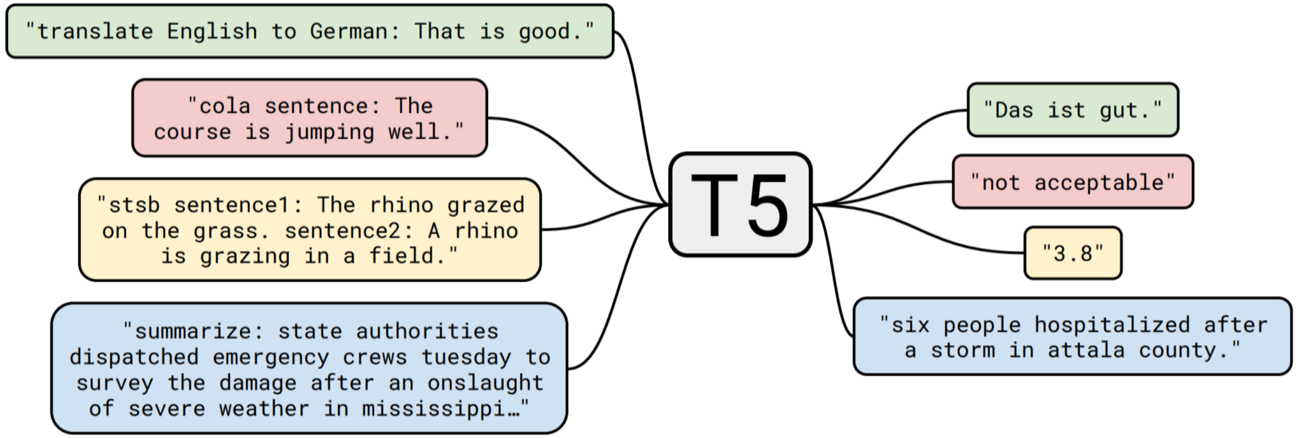
The latest T5 Models are pre-trained on huge cleaned text corpus, about four times the size of Wikipedia

T5 is pre-trained using masked language models to carry out true sentence to sentence prediction

The models are now so advanced, that us mere humans have no need to see the workings out

Models can be up to 11bn parameters off the shelf, which can be fine tuned

There are no token limitations limitations other than compute and memory, and accuracy capabilities when running the models



Practical demonstration – Q&A and document summary on legal text using T5

Aims: Extract data from legal text, for example Contract sign, Parties, Term etc

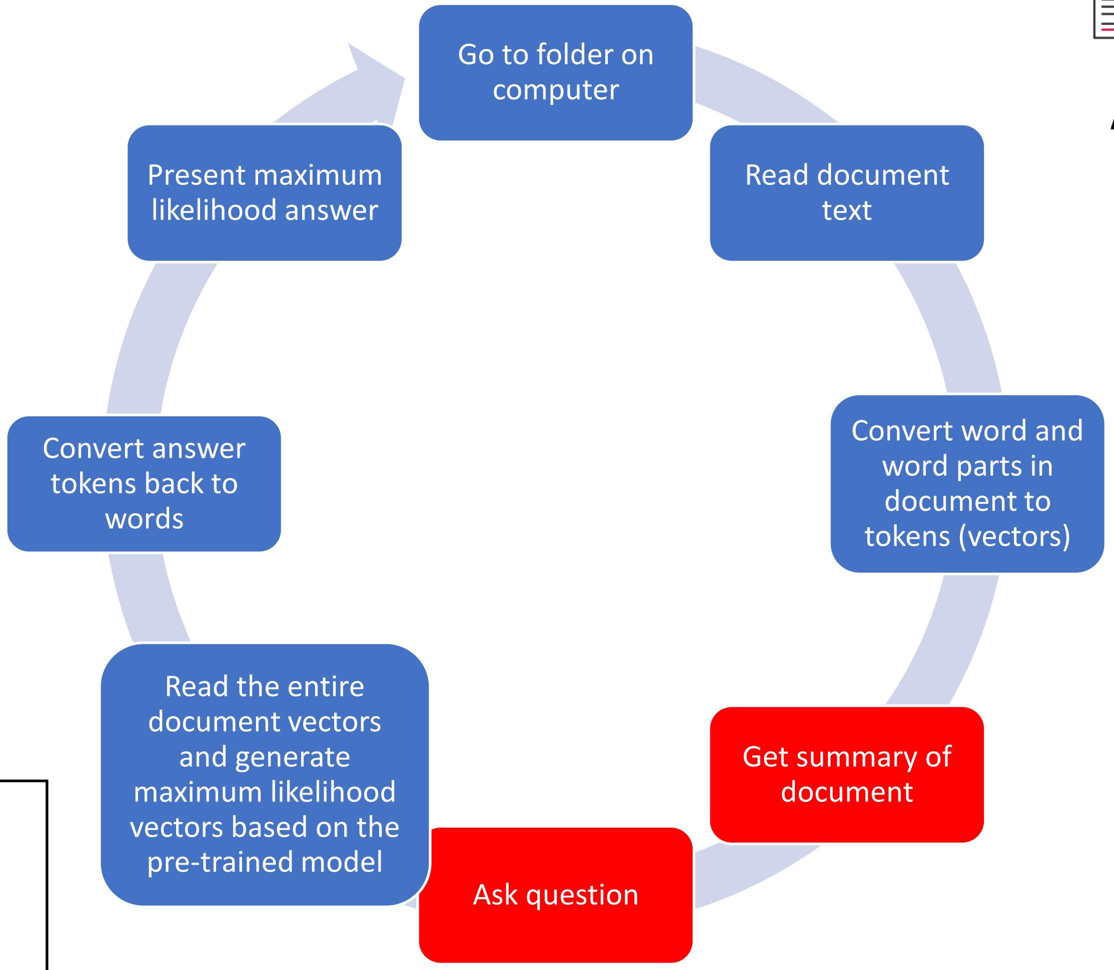
Google have fine tuned the model to address multiple language tasks, however our aims are addressed with SQuAD

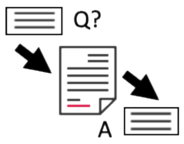
The T5 Large model has approx. 770m parameters, but we can load pre-trained models up to 11bn parameters

Is approximately 85% accurate out of the box

Unlike BERT, this is a true text generalisation model, where we have artificially set the limit to 3000 tokens (10 page legal document)

context	**question**	**answer**
We have found that in the summer, it is generally warm, whereas winter is cold.	When is it warm?	It is warm in the summer





Question Answering

Importance of an AI strategy?

The basics

How to build an AI strategy

Embed AI in your organisation's Toolbox

- Ensure executives and managers understand what machine learning can and can not do, ensure that the knowledge separates hype from reality
- The board or executive, depending on the size of the organisation, should have good conceptual understanding of AI

Get the flywheel spinning

- Your organisation's first AI project should be small enough to be achievable but large enough to be valuable, this will help ensuring the first AI project is successful and build momentum in the organisation.
- The project team needs to include people with different skills from across the organisation, especially those who understand the problem the most (and don't lock them in a room)

The board should be clear on its expectations with regards the project

- Be clear of what problem is being solved, project creep is dangerous.
- Be clear from the start of what the organisation is attempting to optimise, the project timelines including prototyping, MVP, development methodology (eg agile), milestones and what a go/no-go scenario will look like.
- Be clear of Return On Investment metrics and include both skills and infrastructure requirements
- Determine what the impact on stakeholders will be and how this will be communicated
- Consider assurance, reproducibility, bias and potential societal harm

Where to start

- Get a base understanding of what AI, machine learning and deep learning actually is
  - A great starting place is NVIDIA web site, and a short course by deeplearning.ai, called ‘AI for everyone’
- What would you like to do with your data to enhance your business, forgetting the AI
  - Examples: Getting value from your unstructured data, serving enhanced data to clients, document classification, reduce risk, recognise fraud etc
- Do you have the data, and is it in a usable format, to build a model?
  - Example: Corporate data, financial data, ledgers, transaction data, online data sets
  - There are simple tools to get almost any data into a usable format
- What skills do you have to make use of the data to enhance business outcomes
- Are there any immediate quick wins
  - Paper to data, general data extraction

Summary

Artificial Neural Networks are now mainstream

BI and AI are not the same (colouring in versus learning)

You have been using machine learning all of your natural life, so don’t be too surprised when you see how it actually works

The best areas to adopt AI is where it addresses a challenge that is a natural fit in your environment (Don’t look for a problem for AI to solve)

This is not about replacing people, it is about enabling humans to be better at what they do

To ignore AI in 2021 is the same as ignoring the personal computer in the 1980’s. Nobody is using typewriters any more