Climate Finance

Offsets and disclosures

Assignments

Written Project Plan

Due TONIGHT the 23rd by midnight.

Project Plan Presentation

Tuesday the 28th, in class.

Project plan presentation

~3 minutes, 3 slides. (time yourself beforehand to make sure you have the right length!)

Explain the problem, dataset, and planned methods. Include why you've chosen what you have.

Not all team members need to speak in the project plan presentation.

March 28th in class.

ADD YOUR SLIDES TO THE PROVIDED GOOGLE SLIDES PRESENTATION BEFORE CLASS

Climate change in the news

Climate change in the news

POLICY

New Jersey Signs Law To Promote Low-Carbon Concrete With Up To 8% Tax Incentive

by Vasil Velev · February 3, 2023 · 2 minute read

https://carbonherald.com/new-jersey-law-low-carbon-concrete/

The law will come into effect in 2024 and will cover concrete producers that provide a minimum of 50 yards of concrete for any state funded construction projects like bridges, foundation and sidewalks. Producers will be required to prepare and submit an Environmental Product Declaration that includes a score for Global Warming Potential (GWP). The value of GWP will reflect the CO2 generated per 1 cubic meter of concrete.

If the score is below the baseline – something to be established by the New Jersey Department of Environmental Protection before next year – the company will be eligible for a tax up to 8% of total contract cost.

Relevant: **Indianapolis Airport To Use Low Carbon Concrete For Its New Runway**

The New Jersey low-carbon concrete program is capped at an annual budget of $10 million and will be given on a first-come-first-serve basis, with a limit of $1 million per producer.

The law has been two years in the making, with a crucial role from OpenAir Collective volunteers Sue Dorward and Sean Mohan and support from the New Jersey League of Conservation Voters, NRDC, and a broad coalition of environmental and private stakeholders.

Recap

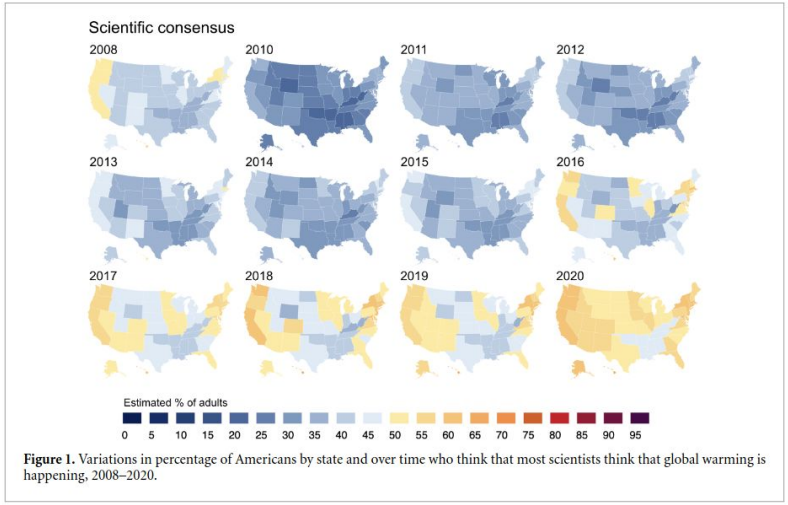
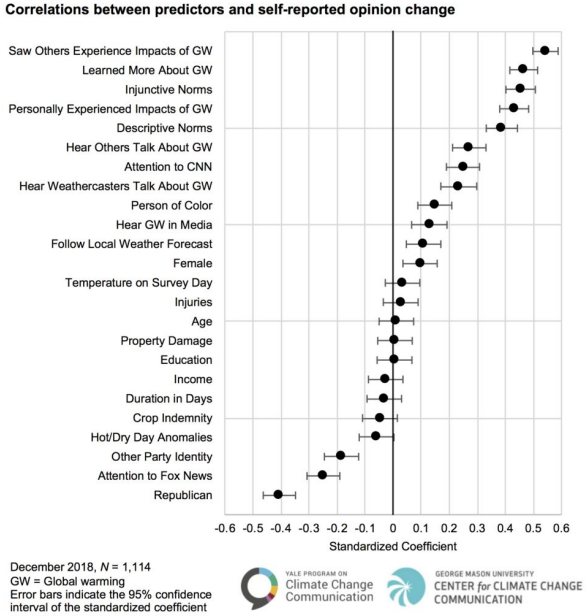
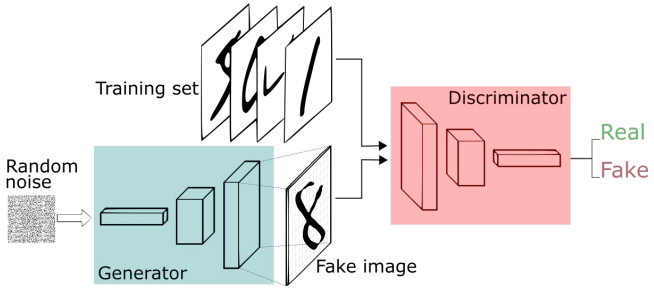


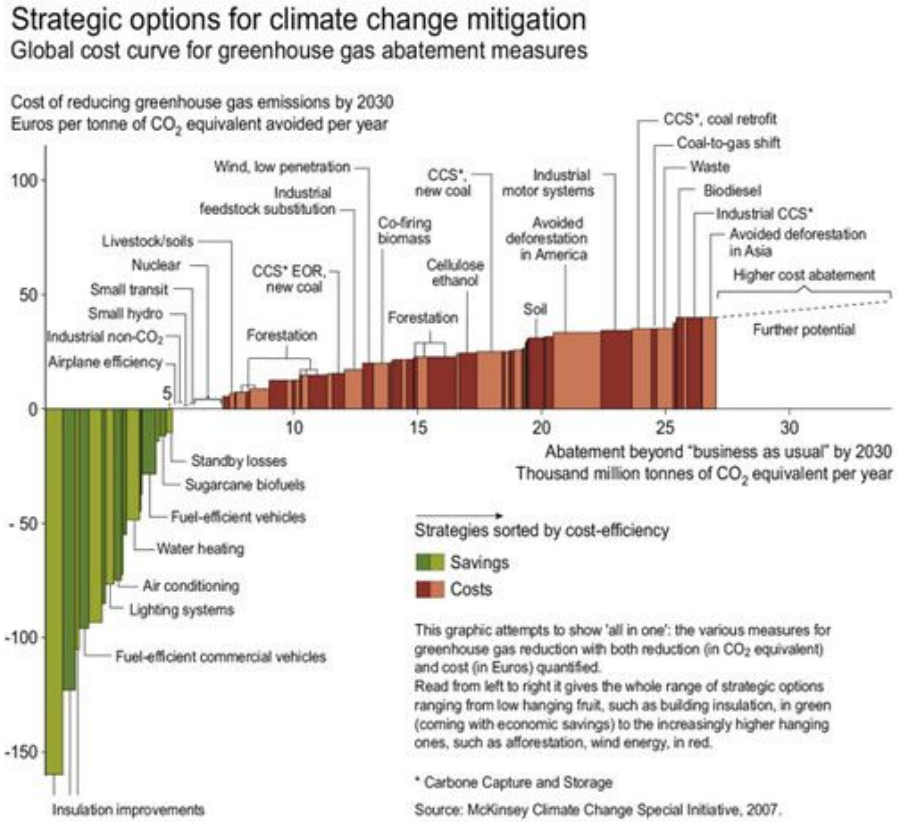
Figure 1. Variations in percentage of Americans by state and over time who think that most scientists think that global warming is happening, 2008–2020.



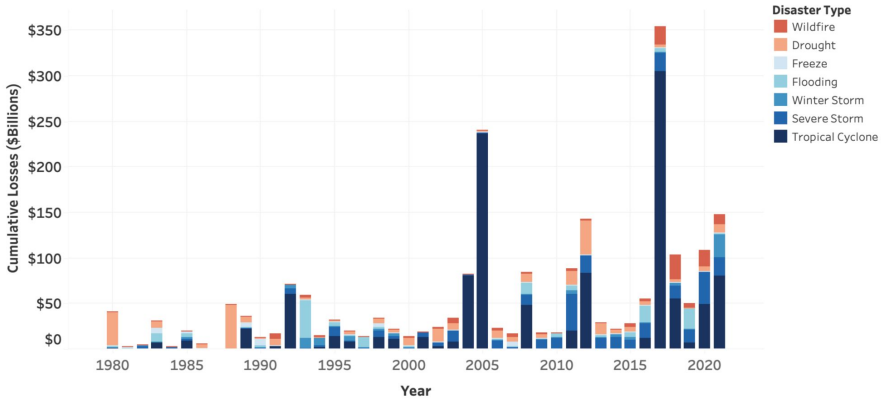


Addressing climate change takes money

Some methods of reducing GHGs are actually financially beneficial (such as increasing energy efficiency), but others will cost a significant amount of money.



But the effects of climate change are even more expensive



$2.2 trillion in losses since 1980 for the US

https://e2.org/reports/cost-of-climate-change/

Press releases

Deloitte Report: Inaction on Climate Change Could Cost the US Economy $14.5 Trillion by 2070

The U.S. economy could gain $3 trillion over the next 50 years if it accelerates towards a path of low-emissions growth

And over the next 50 years, nearly 900,000 jobs could disappear each year due to climate damage

Where should that money come from?

Corporations

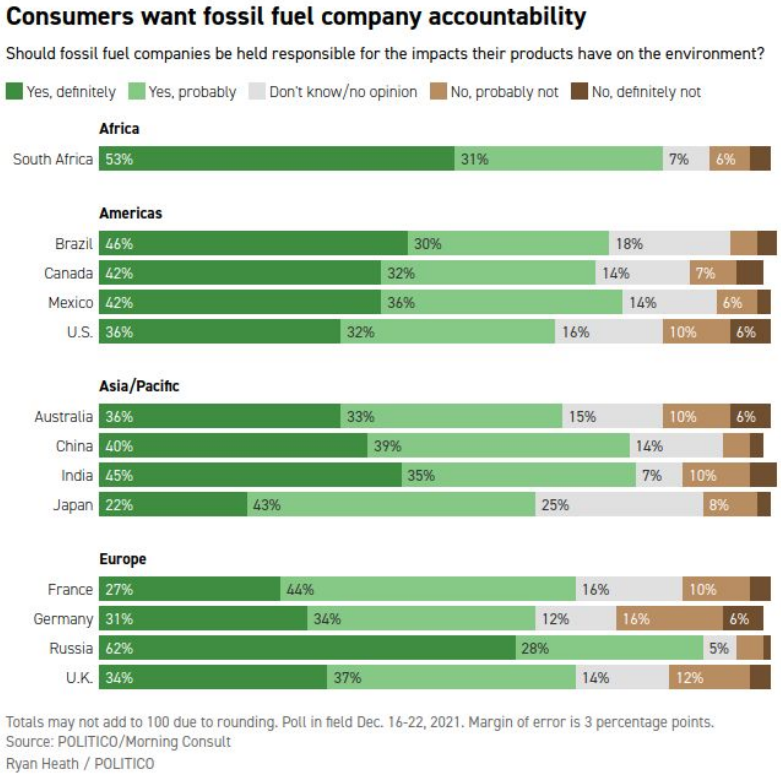
Carbon Tax

Carbon Credits/Offsets

People believe companies should pay

A POLITICO Morning Consult Global Sustainability Poll asked people in 13 countries who should pay — governments, taxpayers, consumers, other countries, or the private sector. In every country but one — India — respondents **singled out companies**.

Respondents in every country surveyed were united **against increasing costs to taxpayers or consumers**. In the U.S., 15 percent of adults said climate change costs should be borne by consumers through higher prices. Eighteen percent said taxpayers should pay a lot of the cost.



https://www.politico.com/news/2022/02/09/climate-change-expensive-who-should-pay-00005198

Taking oil companies to court to pay for climate change

There are at least 20 pending lawsuits filed by cities and states across the U.S., alleging major players in the fossil fuel industry misled the public on climate change to devastating effect.

The claims hinge on longstanding local statutes and common-law torts first widely used in consumer-protection lawsuits from the 1960s and more recently in litigation over tobacco and pharmaceuticals. Key to these laws is that companies can be held accountable for failing to warn consumers of known potential hazards, Sokol said.

In the lawsuits, states and cities are making the case that the fossil fuel industry's failure to warn consumers about its products' contributions to climate change is already having a negative effect on communities.

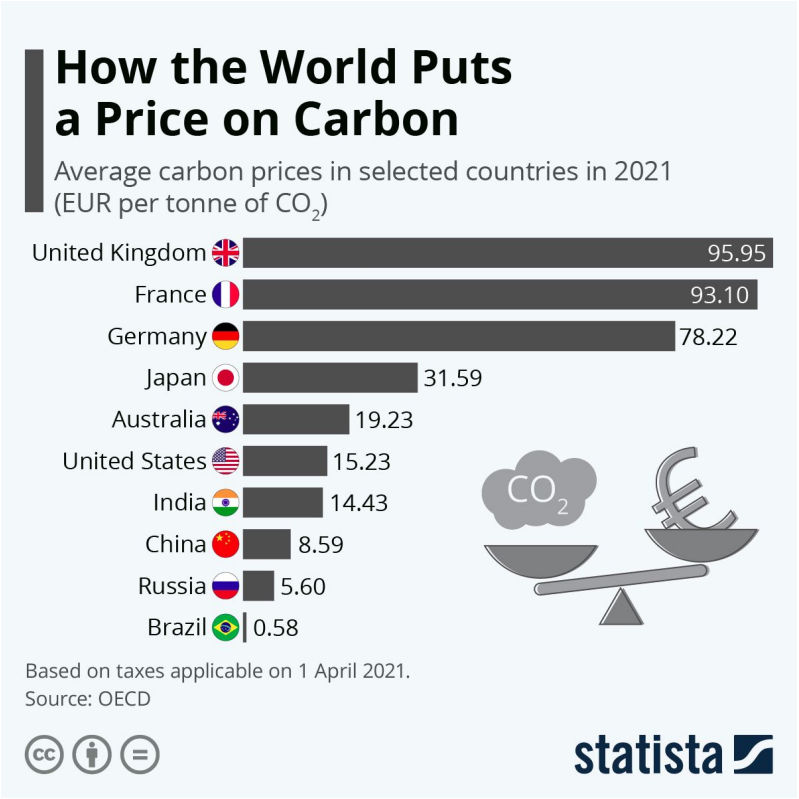
"One thing that might have triggered this wave of litigation is that cities have become aware in the past 15 years that climate change is costing them money," said Doremus, of Berkeley. "That's especially true for coastal cities, counties and states, where a lot of these cases are coming from. I think just looking for any way to deal with this problem has sent them to the state courts."

https://www.pbs.org/wgbh/frontline/article/us-cities-states-sue-big-oil-climate-change-lawsuits/

Carbon tax

According to the EPA, tax-based regulatory systems provide incentives for polluters to find cost-effective solutions to emissions control.

Firms will either pay the tax or, if it is cheaper, they will reduce emissions to avoid the tax.



Offsets and carbon credits

Blockchain-based carbon credit tracking

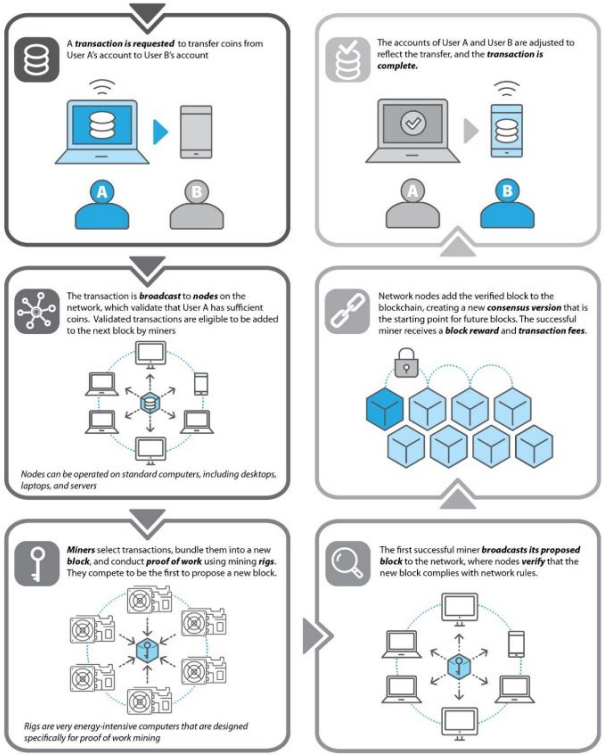


Figure 1.1: Understanding Proof of Work Blockchain in Crypto-Asset Mining. Adapted from Kilroy Blockchain.[33]

The blockchain creates a public ledger that can be used to verify that an entity has bought or sold carbon credits.

Most blockchains run on “proof of work”

Proof of Work requires way too much energy

Any one proposing a blockchain-based carbon credit system or crypto-funded climate company will need to address energy issues

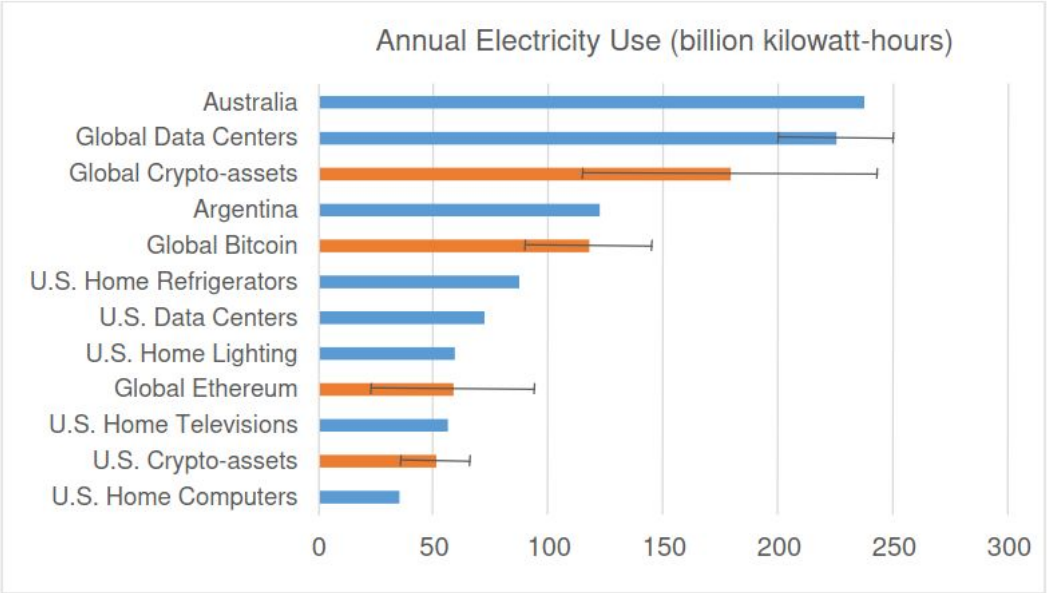
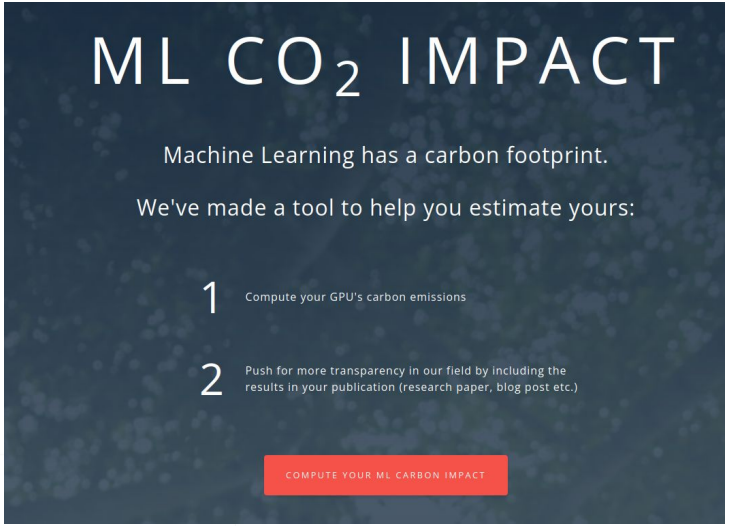


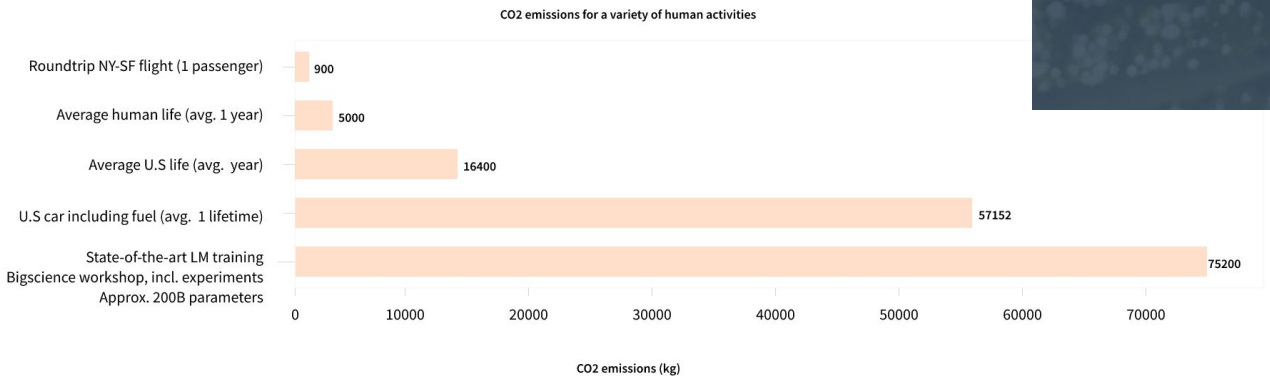
Figure 2.1: Comparison of Annual Electricity Use of Several Examples and the Best Estimates for Crypto-assets, as of August 2022, with error bars representing the best range of values.[80,81]

https://www.whitehouse.gov/wp-content/uploads/2022/09/09-2022-Crypto-Assets-and-Climate-Report.pdf

Machine Learning requires energy too though

Very large models (like ChatGPT) can use an enormous amount of energy for training.





Paper Deep Dive

Analyzing Sustainability Reports Using Natural Language Processing

Alexandra (Sasha) Luccioni
Université de Montréal + Mila

Emily (Emi) Baylor
McGill University

Nicolas Duchene
Université de Montréal

Abstract

Climate change is a far-reaching, global phenomenon that will impact many aspects of our society, including the global stock market [1]. In recent years, companies have increasingly been aiming to both mitigate their environmental impact and adapt to the changing climate context. This is reported via increasingly exhaustive reports, which cover many types of climate risks and exposures under the umbrella of Environmental, Social, and Governance (ESG). However, given this abundance of data, sustainability analysts are obliged to comb through hundreds of pages of reports in order to find relevant information. We leveraged recent progress in Natural Language Processing (NLP) to create a custom model, ClimateQA, which allows the analysis of financial reports in order to identify climate-relevant sections based on a question answering approach. We present this tool and the methodology that we used to develop it in the present article.

https://arxiv.org/pdf/2011.08073.pdf

Background

Climate change will cause a lot of financial impacts.

“It is difficult to predict exactly how and where climate change will impact financial assets, largely due to the **lack of quantitative data** on the subject. ”

In 2019, the Task Force on Climate-related Financial Disclosures estimated that out of the 1000 companies whose reports they analyzed, only 29% made relevant climate disclosures, stating that they were “***concerned that not enough companies are disclosing decision-useful climate-related financial information***" [4].

“This data is often in textual format, buried in hundreds of pages of financial documents which must be manually analyzed, **requiring significant time and effort**.

The goal

Create a tool allowing more efficient analysis of financial reports, reducing the time and effort required to identify climate-relevant disclosures.

Specifically, find a sentence in a financial report that answers a specific climate disclosure question.

Brainstorm

What kind of data would you want to have to be able to approach this problem?

What kind of methods would you apply?

How would you measure success?

What difficulties might you face?

Natural Language Processing

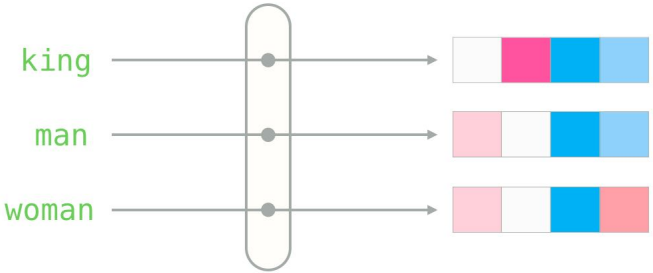
NLP requires building algorithms that can make sense of text.

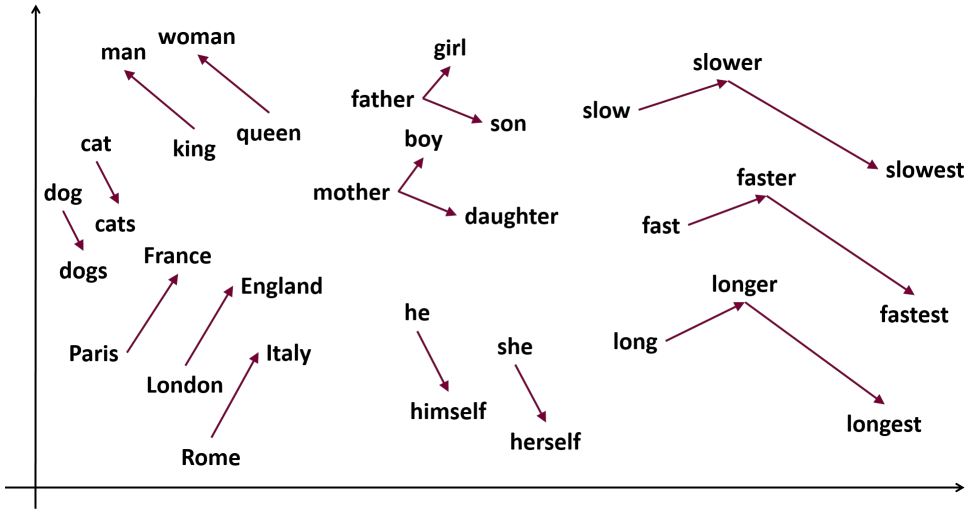
NLP tasks can be incredibly challenging due to the diverse ways in which people use language and how language relates to the real world.

https://datasciencedojo.com/blog/natural-language-processing-applications/

Natural Language Processing

Requirement: Represent meaning as a vector of numbers





Natural Language Processing

Requirement: Represent meaning as a vector of numbers

Simplest approach = represent words in terms of how often they co-occur with other words.

	Roses	are	red	Sky	is	blue
Roses	1	1	1	0	0	0
are	1	1	1	0	0	0
red	1	1	1	0	0	0
Sky	0	0	0	1	1	1
is	0	0	0	1	1	1
Blue	0	0	0	1	1	1

This paper’s approach

Use a “Large Language Model” (LLM)

Feed it a question and sentence from the financial report and train it to determine if they are a match or not

Large Language Model lineage

RoBERTa: “Robustly optimized **BERT** approach”

BERT: “Bidirectional Encoder Representations from **Transformers**”

Transformers:

Attention Is All You Need

Ashish Vaswani[*] Google Brain avaswani@google.com

Noam Shazeer[*] Google Brain noam@google.com

Niki Parmar[*] Google Research nikip@google.com

Jakob Uszkoreit[*] Google Research usz@google.com

Llion Jones[*] Google Research llion@google.com

Aidan N. Gomez[*] [†] University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser[*] Google Brain lukaszkaiser@google.com

Illia Polosukhin[*] [‡] illia.polosukhin@gmail.com

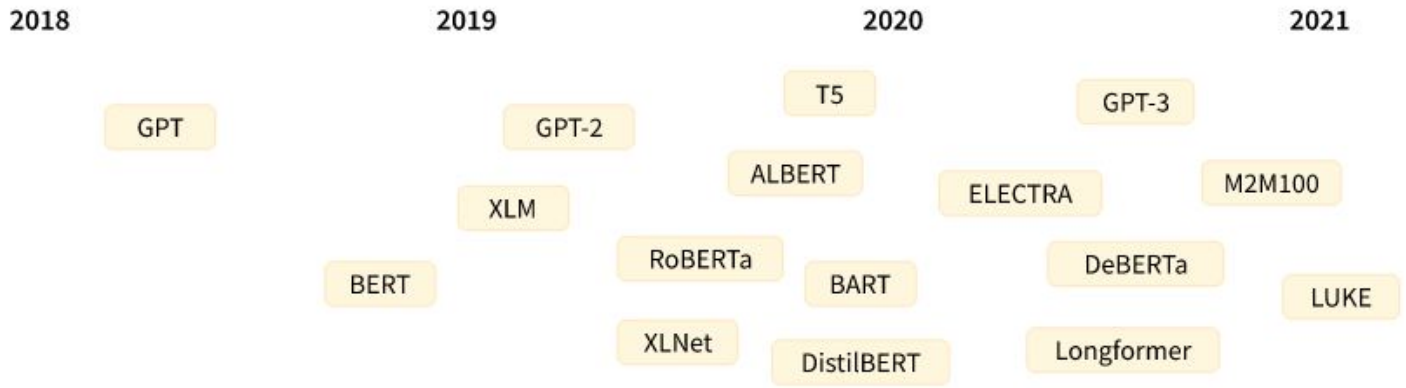
Abstract

The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.0 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature.

“self-attention”

Large Language Model lineage

Here are some reference points in the (short) history of Transformer models:

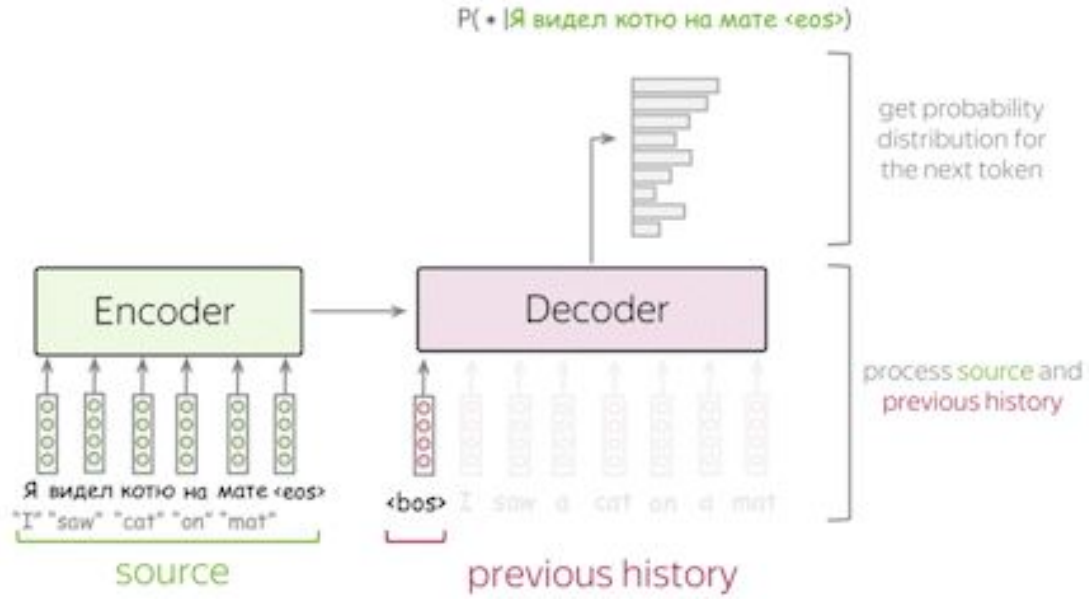


The Transformer architecture was introduced in June 2017. The focus of the original research was on translation tasks. This was followed by the introduction of several influential models, including:

https://huggingface.co/course/chapter1/4?fw=pt

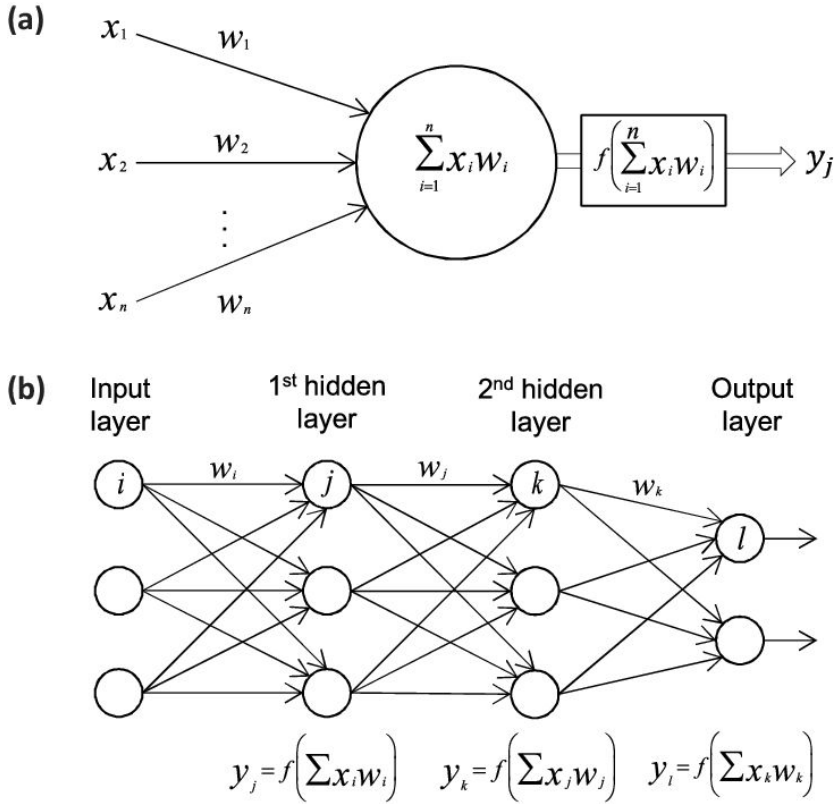
Architecture of a Large Language Model

Many language tasks are “sequence to sequence” problems that can be solved with an encoder and decoder. The encoder and decoder are each artificial neural networks

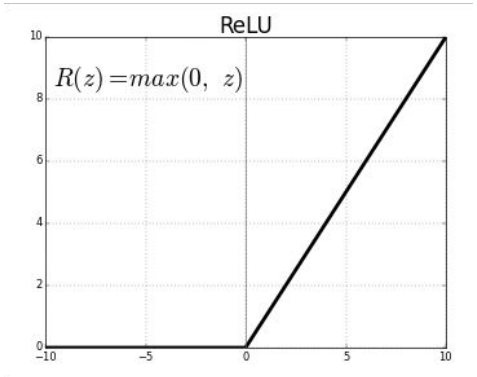


https://lena-voita.github.io/nlp_course/seq2seq_and_attention.html

Neural networks

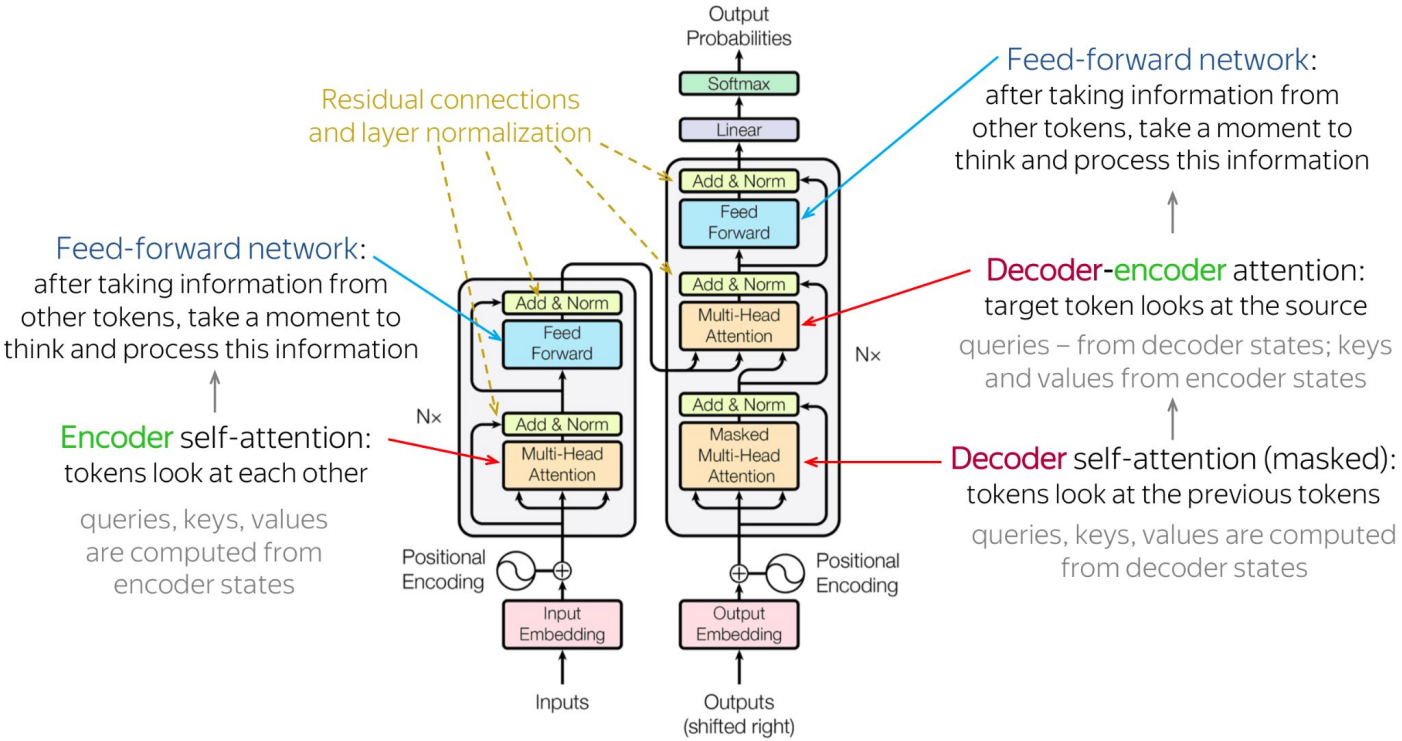


Basic or “vanilla” networks multiple weights by node activity, sum these values, and rectify the sum.



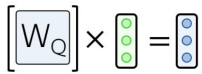
Vieira et al.

Transformer architecture



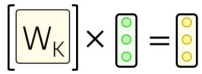
Transformer architecture

Each vector receives three representations (“roles”)



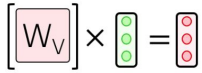
Query: vector **from** which the attention is looking

“Hey there, do you have this information?”



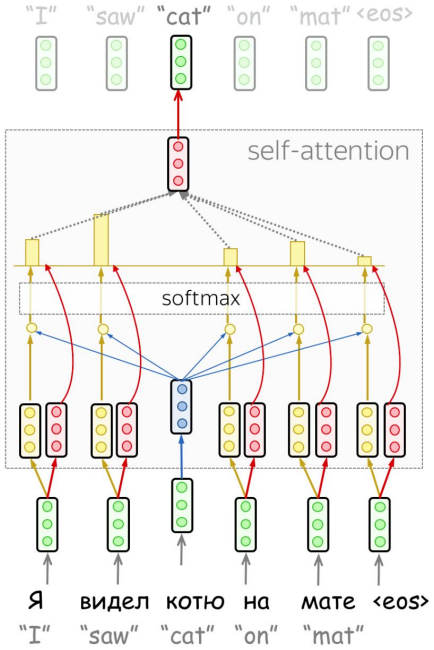
Key: vector **at** which the query looks to compute weights

“Hi, I have this information – give me a large weight!”



Value: their weighted sum is attention output

“Here’s the information I have!”



Key insight: combine information across words. This is known as “self-attention”.

I arrived at the bank after crossing the … …street? …river?

What does **bank** mean in this sentence?

RNNs

O(N) steps to process a sentence with length N

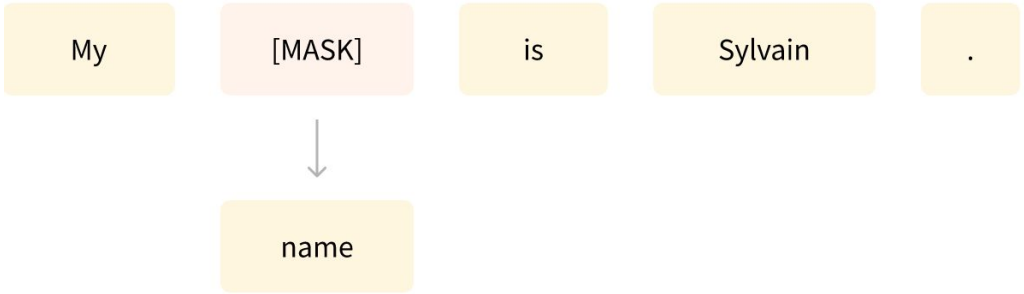
Transformer

Constant number of steps to process any sentence

Training RoBERTa

Not trained on language translation or sentence generation.

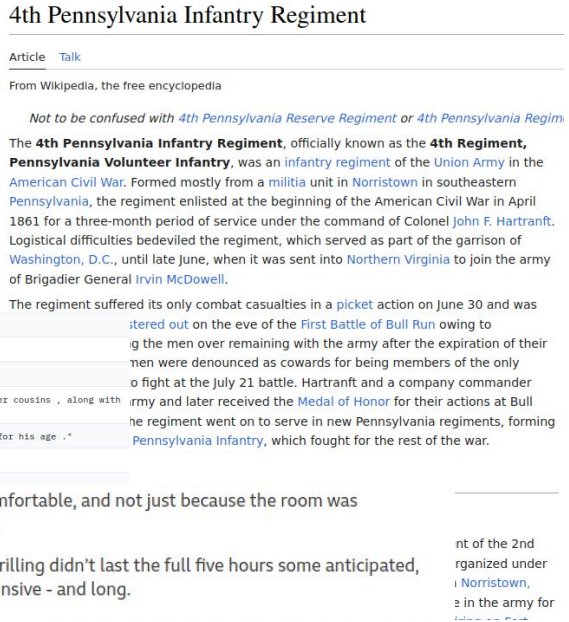
Trained with a “masking” task: predict hidden word.



The data

Pre-trained on a very large amount of text:

160 GB of text from Books Corpus, English Wikipedia, CommonCrawl News dataset (63 million articles), Web text corpus, and Stories from Common Crawl.


4th Pennsylvania Infantry Regiment

Article Talk

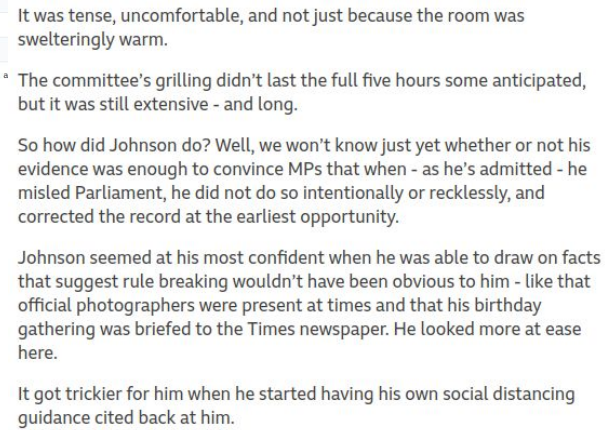
From Wikipedia, the free encyclopedia

Not to be confused with 4th Pennsylvania Reserve Regiment or 4th Pennsylvania Regim

The **4th Pennsylvania Infantry Regiment**, officially known as the **4th Regiment, Pennsylvania Volunteer Infantry**, was an infantry regiment of the Union Army in the American Civil War. Formed mostly from a militia unit in Norristown in southeastern Pennsylvania, the regiment enlisted at the beginning of the American Civil War in April 1861 for a three-month period of service under the command of Colonel John F. Hartranft. Logistical difficulties bedeviled the regiment, which served as part of the garrison of Washington, D.C., until late June, when it was sent into Northern Virginia to join the army of Brigadier General Irvin McDowell.

The regiment suffered its only combat casualties in a picket action on June 30 and was stered out on the eve of the First Battle of Bull Run owing to g the men over remaining with the army after the expiration of their nen were denounced as cowards for being members of the only o fight at the July 21 battle. Hartranft and a company commander rmy and later received the Medal of Honor for their actions at Bull he regiment went on to serve in new Pennsylvania regiments, forming Pennsylvania Infantry, which fought for the rest of the war.

nt of the 2nd rganized under Norristown, e in the army for


It was tense, uncomfortable, and not just because the room was swelteringly warm.

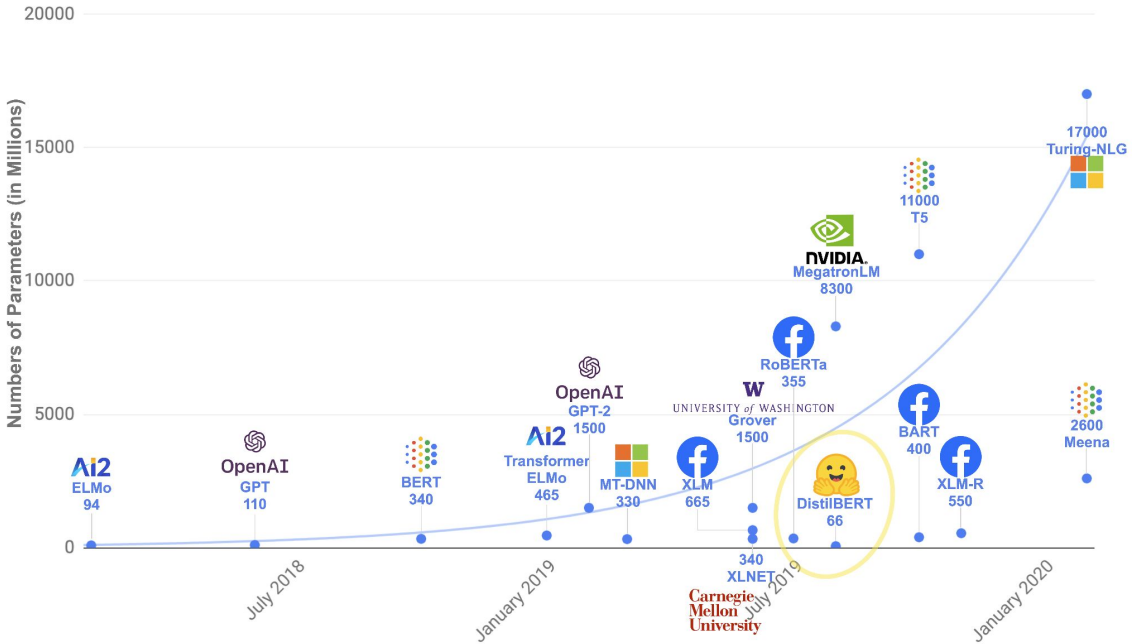
The committee's grilling didn't last the full five hours some anticipated, but it was still extensive - and long.

So how did Johnson do? Well, we won't know just yet whether or not his evidence was enough to convince MPs that when - as he's admitted - he misled Parliament, he did not do so intentionally or recklessly, and corrected the record at the earliest opportunity.

Johnson seemed at his most confident when he was able to draw on facts that suggest rule breaking wouldn't have been obvious to him - like that official photographers were present at times and that his birthday gathering was briefed to the Times newspaper. He looked more at ease here.

It got trickier for him when he started having his own social distancing guidance cited back at him.

These models are very large



These models can take days or weeks to train on large GPU clusters (and therefore use a lot of energy).

Therefore, individual research groups do not train their own. They simply download the connection weights of pre-trained models.

Data problem!

Progress in NLP applications in finance has proven to be challenging notably because of the specialized language used: terms such as 'bull' and 'short' do not have the same meanings in finance as in general discourse [1], whereas technical terms such as 'liquidity' or 'Keynesian' may not even be present in training corpora. In fact, research in financial NLP has found that using general-purpose NLP models trained on corpora such as Wikipedia and the Common Crawl fail to capture domain-specific terms and concepts which are critical for a coherent representation of the financial lexicon, and are therefore difficult to use out-of-the-box for financial tasks [14].

The text this model has been trained on doesn't contain enough financial terms.

What can be done?

Fine tune on financial reports

2,249 publicly available financial and sustainability reports pulled from sources such as the Securities and Exchange Commission and the Global Reporting Initiative databases.

Train the model a bit more with these text sources.

Fine tune on financial reports

2,249 publicly available financial and sustainability reports pulled from sources such as the Securities and Exchange Commission and the Global Reporting Initiative databases.

Train the model a bit more with these text sources.

***Annoying technical issue*:**

“We extracted the raw text from the PDFs of the reports...sentences that were part of a PDF table were often not identified by our approach”

Training procedure

1. Use a large language model* that was pre-trained on Wikipedia and Common Crawl

2. Fine tune on financial documents

*In this paper, they tested two versions of RoBERTa, one with 125m parameters and one with 355m. They found the larger one didn’t increase performance enough to warrant the increased compute costs

Training procedure

1. Use a large language model* that was pre-trained on Wikipedia and Common Crawl

2. Fine tune on financial documents

3. Supervised training on labeled question data

The data

“We reached out to a team of sustainability analysts, who were able to provide us with a small set of financial reports from previous years, hand-labeled using the 14 questions from the Task Force on Climate-related Financial Disclosures.”

TCFD Question
1) Does the organization describe the board’s oversight of climate-related risks and / or opportunities?
2) Does the organization describe management’s role in assessing and managing climate-related risks and/or opportunities?
3) Does the organization describe the climate-related risks or opportunities the organization has identified?
4) Does the organization describe time frames (short, medium, or long term) associated with its climate-related risks or opportunities?
5) Does the organization describe the impact of climate-related risks and opportunities on the organization?
6) Does the organization describe the resilience of its strategy, taking into consideration different climate-related scenarios, including a potential future state aligned with the Paris Agreement?
7) Does the organization disclose the use of a 2C scenario in evaluating strategy or financial planning, or for other business purposes?
8) Does the organization describe the organization’s processes for identifying and/or assessing climate-related risks?
9) Does the organization describe the organization’s processes for managing climate-related risks?
10) Does the organization describe how processes for identifying, assessing, and managing climate-related risks are integrated into the organization’s overall risk management?
11) Does the organization disclose the metrics it uses to assess climate- related risks and/or opportunities?”
12) Does the organization disclose Scope 1 and Scope 2, and, if appropriate Scope 3 GHG emissions?
13) Does the organization describe the targets it uses to manage climate- related risks and/or opportunities?
14) Does the organization describe its performance related to those targets (referenced in question 13)?

The data

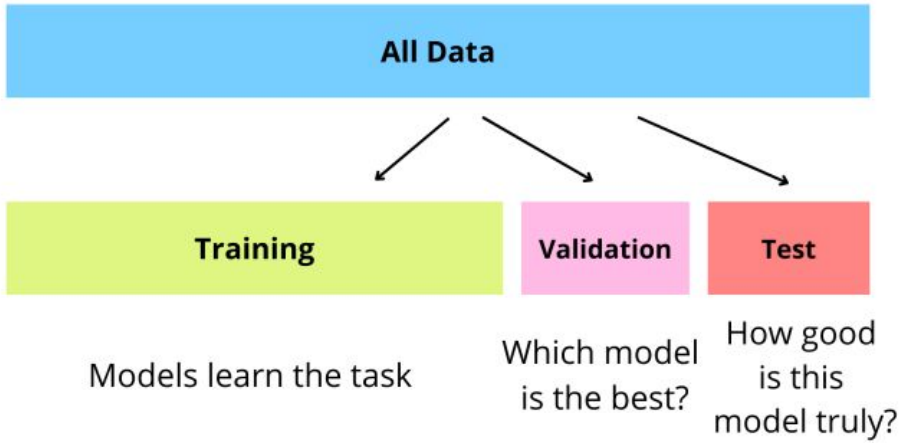
“Positive examples consisted of pairs of questions and sentences which contained the answers to the questions, whereas negative examples were generated by pairing the remaining sentences with the questions that they did not answer”

Table 1: Examples of Question-Answer pairs from our corpus

TCFD Question	**Answer Passage**
Does the organization describe the board’s (or board committee’s) oversight of climate-related risks and/or opportunities?	*The Company’s Audit Committee has the delegated risk management oversight responsibility and receives updates on the risk management processes and key risk factors on a quarterly basis.*
Does the organization describe the climate-related risks or opportunities the organization has identified?	*The availability and price of these commodities are subject to factors such as changes in weather conditions, plantings, and government policies*

“Our ClimateQA model was trained on 15,000 negative examples and 1,500 positive examples, whereas the development set comprised of 7,500 negative examples and 750 positive examples, while the test set had 1,200 negative and 400 positive examples”

“Development” or “validation” data



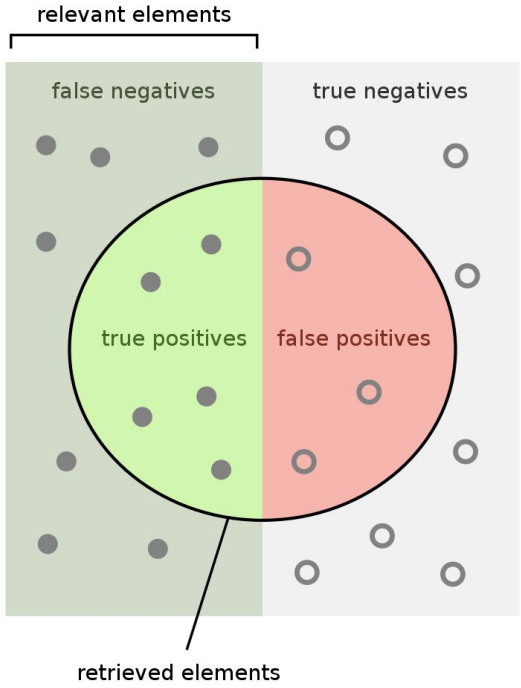
Evaluation

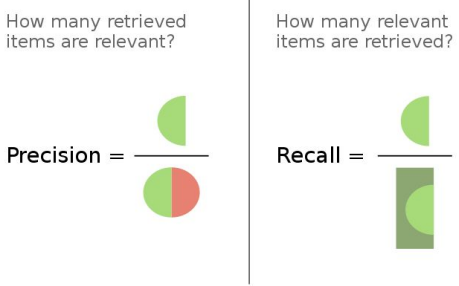
Evaluation

Binary classification with imbalanced data

$$F_1 = \frac{2}{\mathrm{recall}^{-1} + \mathrm{precision}^{-1}} = 2\frac{\mathrm{precision} \cdot \mathrm{recall}}{\mathrm{precision} + \mathrm{recall}} = \frac{2\mathrm{tp}}{2\mathrm{tp} + \mathrm{fp} + \mathrm{fn}}.$$

F1 ranges from 0 to 1





Results

Results

Table 3: ClimateQA results by sector

	Validation F1 Score	Test F1 Score	Val - Test Difference
Agriculture, Food & Forests	89.4%	72.1%	-17.2%
Energy	94.2%	89.8%	-4.4%
Banks	91.9%	86.6%	-5.3%
Transportation	86.9%	72.5%	-14.4%
Insurance	92.9%	78.7%	-14.2%
Materials & Buildings	91.8%	67.6%	-24.2%
Average across sectors	**91.7%**	**82.0%**	**-9.7%**

“We found that the Energy sector had the best results, most likely due to the homogeneity of the companies in the labeled data we received – most of it was from oil and gas companies who disclosed very similar risks and opportunities and often used extensive boilerplate language.”

Implementation

The model must be easily available and simple to interact with in order for it be used and have impact.

“We have spent a significant amount of time and effort deploying our ClimateQA model. To this end, the model is hosted on the Microsoft Azure cloud, allowing users to interact with a web application without needing ML expertise.

Via the website, a user is able to upload PDF files to be analyzed and receive a batch ID which they can subsequently use to check if they have been processed.”

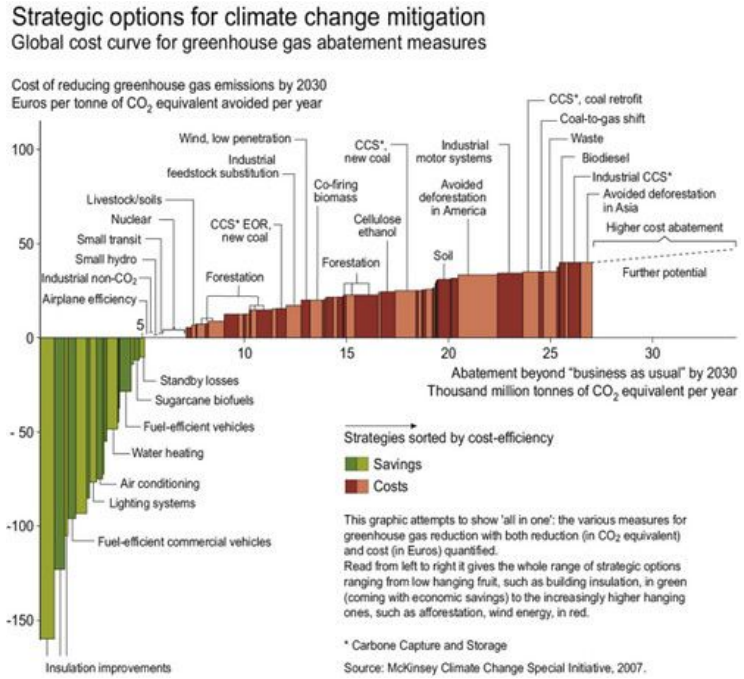
Further Resources

Review of Machine Learning methods used in Climate Finance: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4352569

Microsoft report on offsets: https://www.microsoft.com/en-us/corporate-responsibility/sustainability/carbon-removal-program

NLP online course: https://lena-voita.github.io/nlp_course

Summary



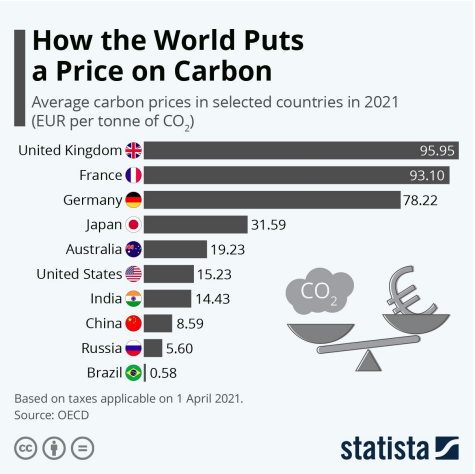


Table 1: Examples of Question-Answer pairs from our corpus

TCFD Question	Answer Passage
Does the organization describe the board's (or board committee's) oversight of climate-related risks and/or opportunities?	*The Company's Audit Committee has the delegated risk management oversight responsibility and receives updates on the risk management processes and key risk factors on a quarterly basis.*
Does the organization describe the climate-related risks or opportunities the organization has identified?	*The availability and price of these commodities are subject to factors such as changes in weather conditions, plantings, and government policies*