


The background features several green circles and leaves. One leaf in the upper right shows a grid pattern with numbers. Another leaf in the lower left shows a close-up of its veins. A third leaf in the lower center shows a grid pattern with numbers.

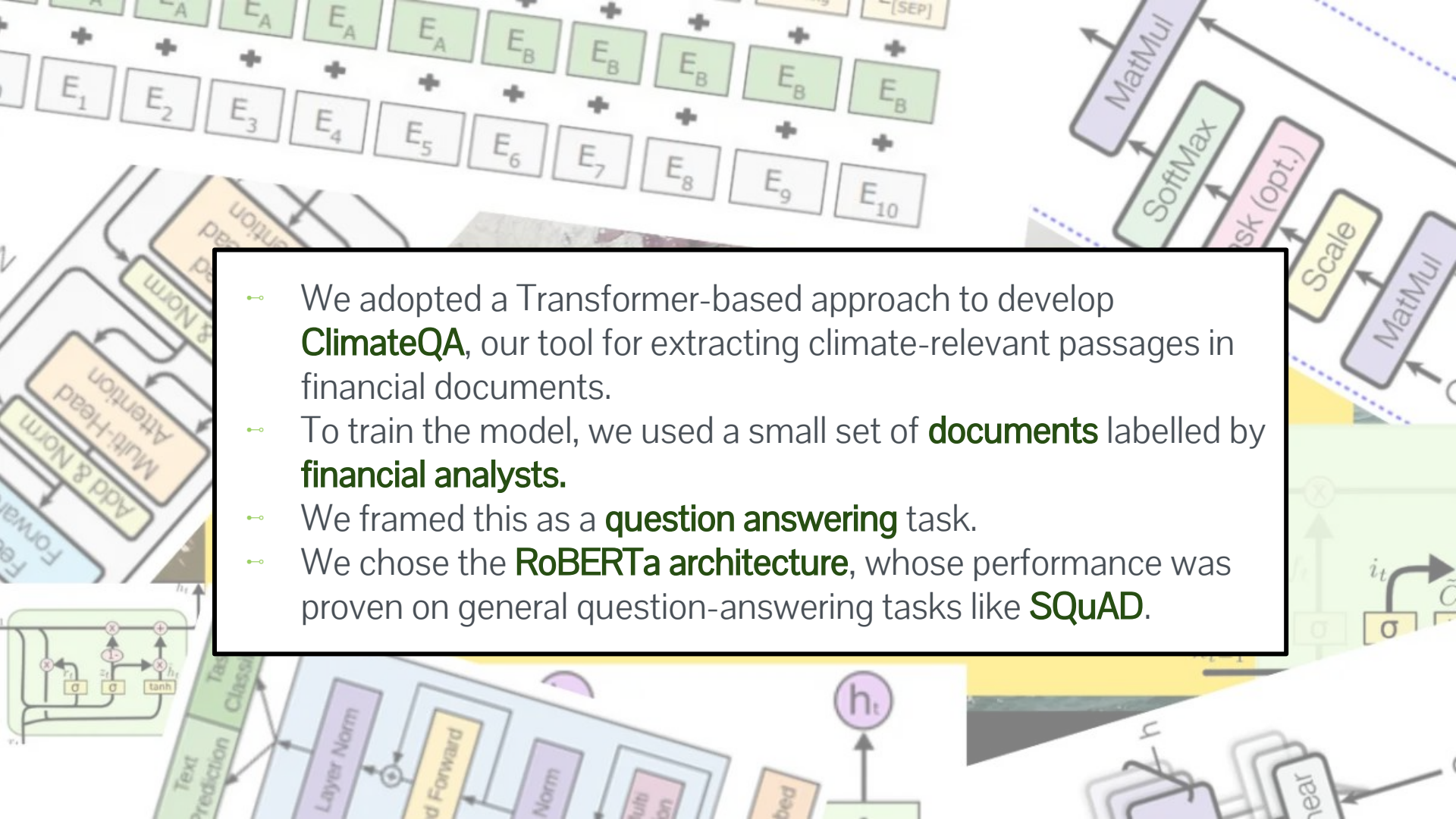
# Analyzing Sustainability Reports Using Natural Language Processing

Sasha Luccioni, Emi Baylor, Nicolas Duchene

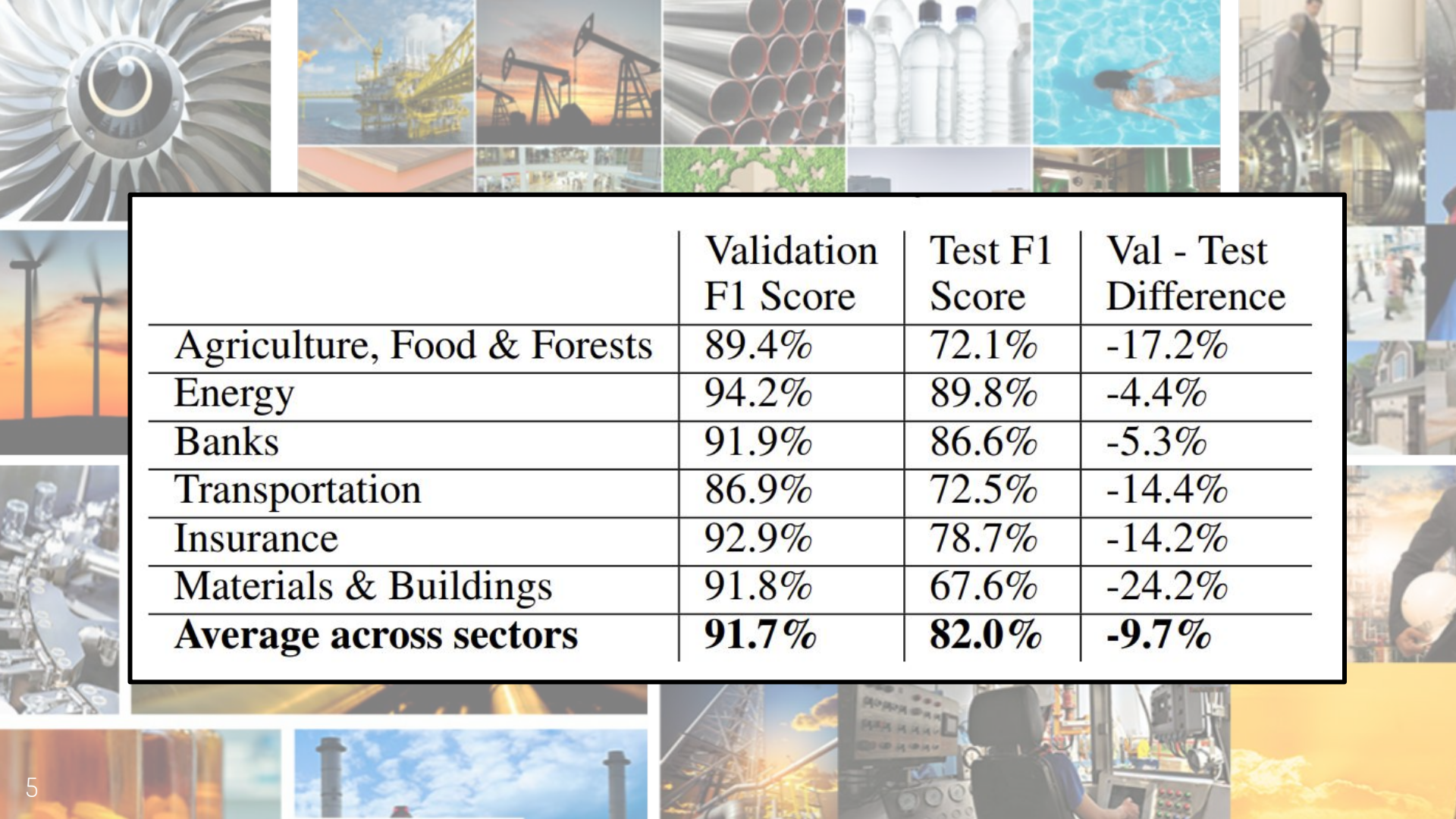
*Université de Montréal, McGill University, Mila*

- 
- Climate change poses a substantial risk to global assets and stocks, measured in the **trillions of dollars**.
  - It is hard to forecast **where, how, or when** climate change will impact financial assets, largely due to the lack of quantitative data on the subject.
  - Gathering data regarding the **risks** and **exposure** that climate change poses to specific companies is a key part of predicting the extent of climate change impacts on the stock market.

- Disclosing climate change **risks** and **liabilities** currently consists of a mix of mandatory and voluntary initiatives.
- In 2015, the **Task Force on Climate-related Financial Disclosures (TCFD)** was founded to improve the state of voluntary climate disclosing and to encourage companies to increase their climate transparency.
- They released a set of **14 questions** to guide sustainability reporting.
  - These questions are extensively used to guide the analysis of climate risk disclosures, with analysts using them to assess the extent and type of **climate exposure** of companies.

- 
- The background features a collage of various neural network diagrams. At the top, there's a sequence of boxes labeled E1 through E10, with some in green (EA, EB) and some in white (E1-4, E6-10), connected by plus signs. To the right, a vertical stack of boxes includes 'MatMul' (purple), 'SoftMax' (green), 'Task (opt.)' (pink), 'Scale' (yellow), and another 'MatMul' (purple). Below this, a diagram shows a 'Forward' pass with 'Add & Norm' and 'Multi-Head Attention' blocks. At the bottom, a diagram shows a 'Text Prediction' task with a 'Layer Norm' block, a 'Forward' block, and a 'Norm' block, leading to a hidden state 'hi'.
- We adopted a Transformer-based approach to develop **ClimateQA**, our tool for extracting climate-relevant passages in financial documents.
  - To train the model, we used a small set of **documents** labelled by **financial analysts**.
  - We framed this as a **question answering** task.
  - We chose the **RoBERTa architecture**, whose performance was proven on general question-answering tasks like **SQuAD**.





	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%
<b>Average across sectors</b>	<b>91.7%</b>	<b>82.0%</b>	<b>-9.7%</b>

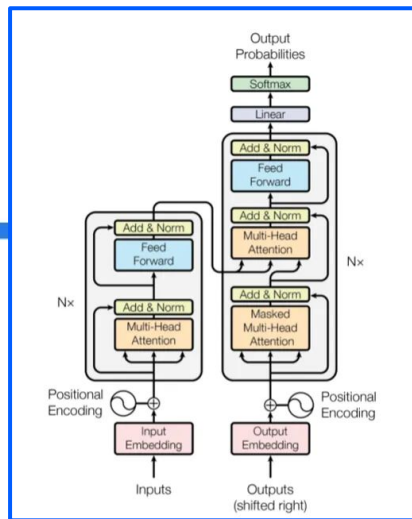
## TCFD Question

Does the organization describe the climate-related risks or opportunities the organization has identified?

## Relevant Passages

*We also understand there can be a financial impact on our operations from climate-related risks.*

*We continue to develop processes to quantify the potential financial impacts of climate-related risks and the costs of actions taken to manage these risks.*



ClimateQA

# Thank you!

Come see us at the  
poster session!

