

How AI technology is used by researchers and companies in R&D processes

Dr Rachel Free



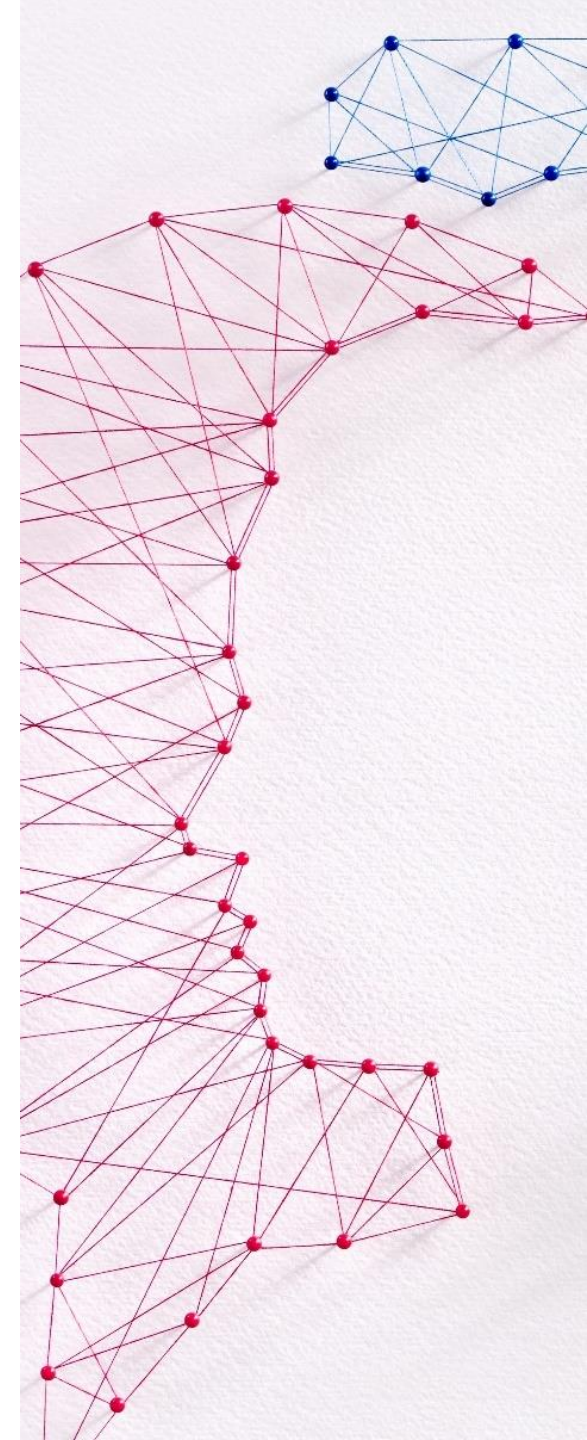
What we will cover today

Client using GPT-4o

Fictional invention creation scenario in the computer technology field

Using GPT-4o

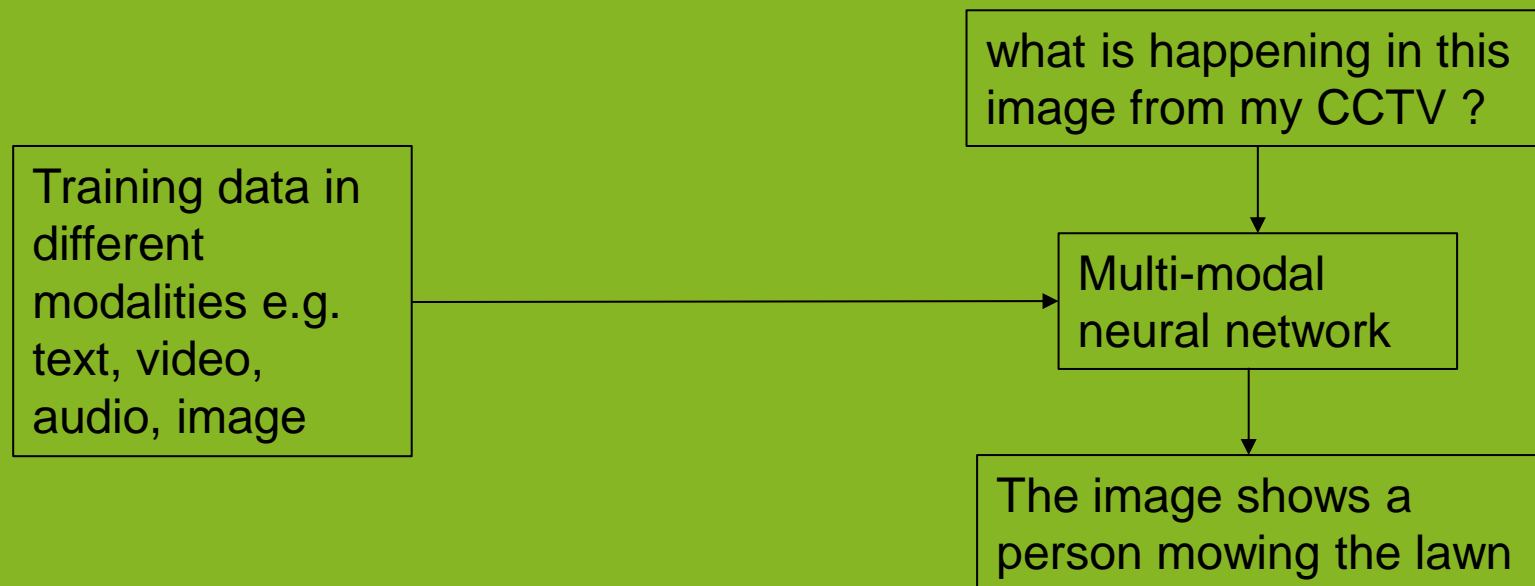
- Enquiry from a client
- Client had an idea for a new service
- Client used GPT-4o to provide the new service by using GPT-4o to analyse data and compute forecasts
- In this example, GPT-4o is being used as a tool to analyse data; the idea for the new service itself was conceived by the human
- The human could be an inventor in jurisdictions where thinking of a problem is enough to be an inventor.



Fictional example

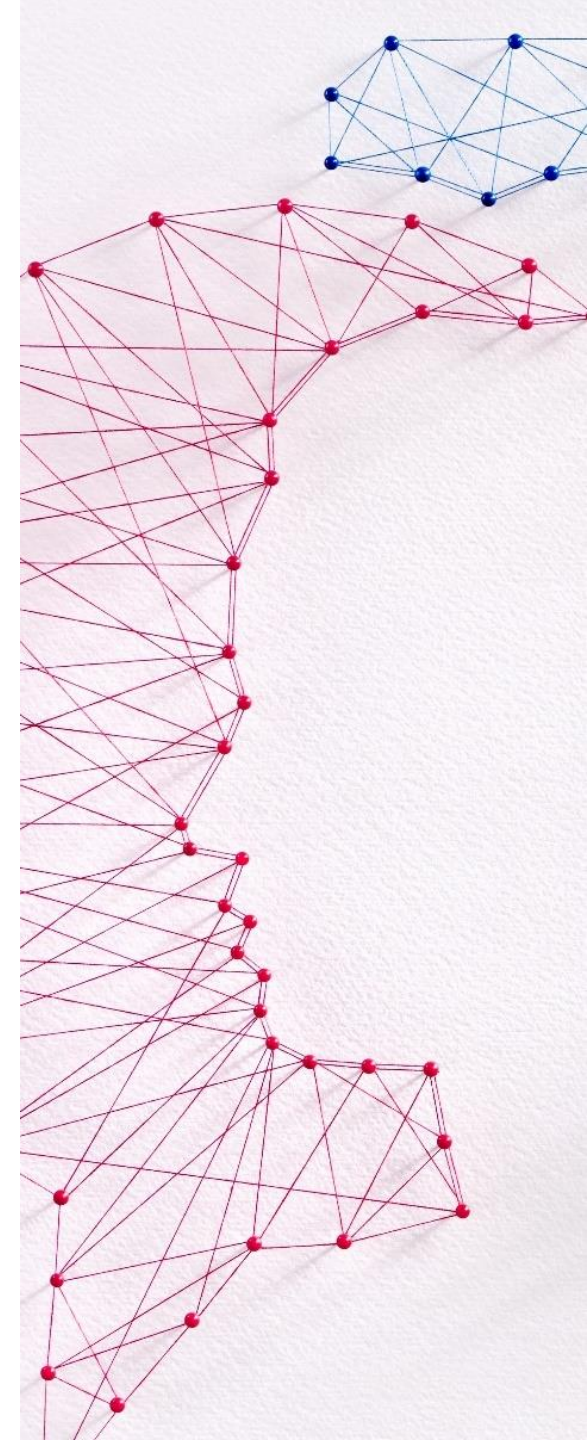


Manager assigns the research and development team a new task: make a better multi-modal model



Review existing multi-modal models

- The team carries out a **review of existing multi-modal models** and finds that they can explain images reasonably well and that text prompts can be given to generate images reasonably well
- The team **wonders what the manager means** by “a better multi-model model”
- Alex realises that existing multi-modal models are **not very good at integrating information across modalities and they can’t generate arbitrary sequences of images and text**

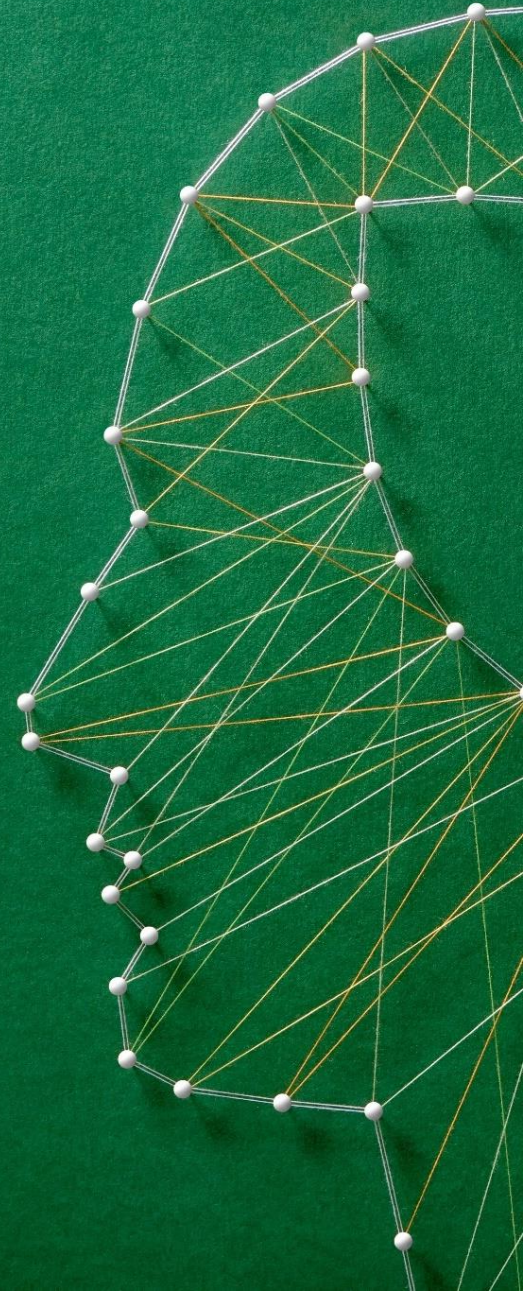


Weekly team meeting

In the weekly team meeting, the team decides to make a multi-modal model that is good at integrating information across modalities and can generate arbitrary sequences of images and text

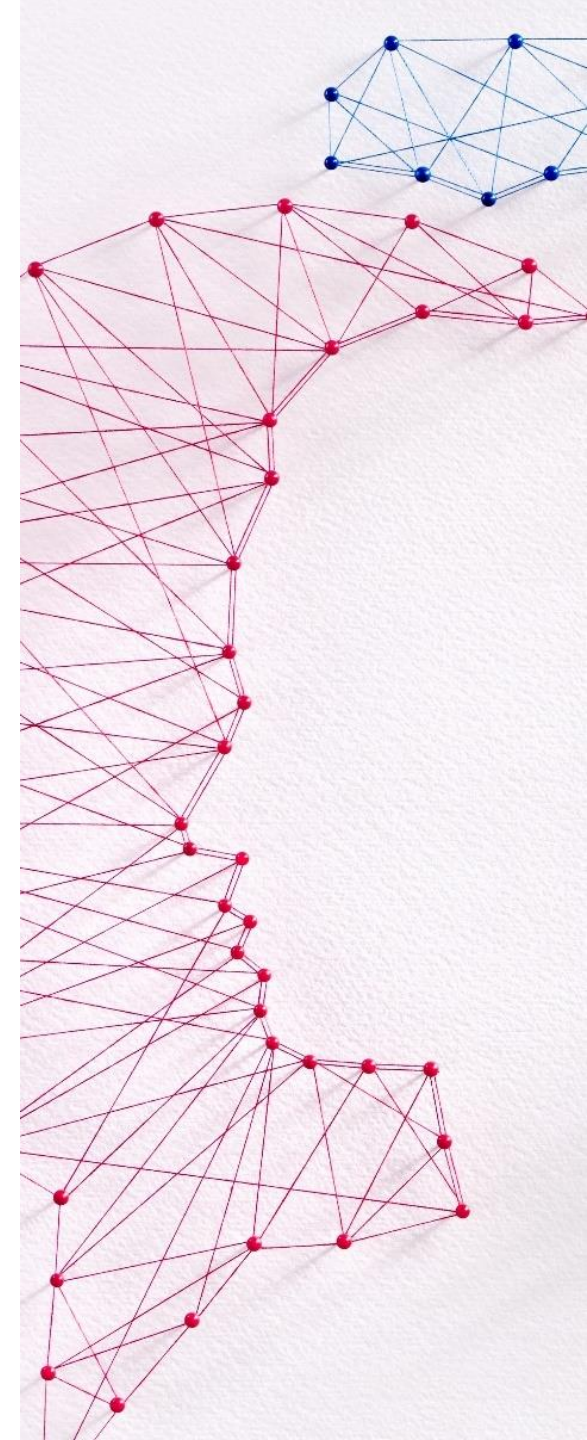
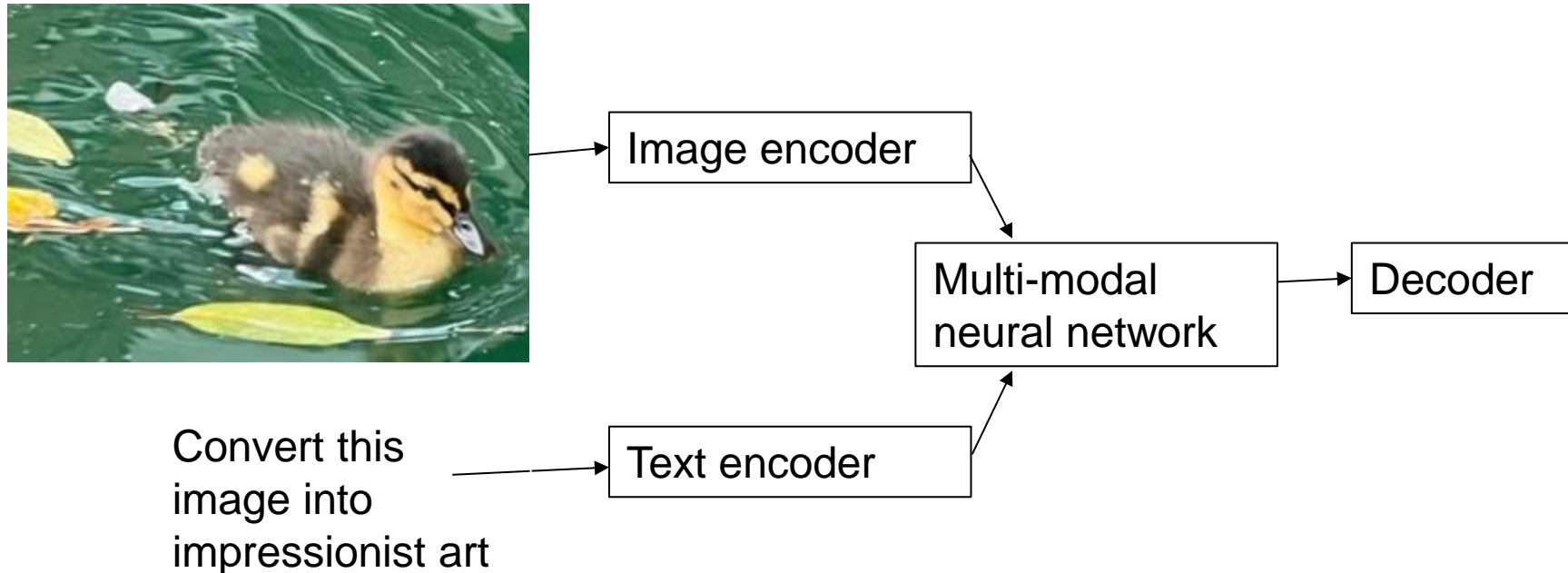
Sam says that the best solution would be to **increase the number of parameters** in the model since that has been a way to obtain “emergent behaviours” from neural networks that generate text

Jamie says that it would be a good idea to obtain more training data perhaps by making some synthetic training data from different modalities



Theoretical insight

George reads research papers about existing multi-modal models and notices that the images are encoded using a separate encoder from the text encoder. George suggests the team try to make an encoder that can encode both text and images in case it will enable the model to achieve the team's goal.





The team decides to follow George's idea as it seems the most theoretically sensible idea

**AI TOOLS CAN BE USED TO
PRIORITIZE WHICH IDEAS TO
TRY FIRST**



CMS
law·tax·future



The whole team work together to make a multi-modal model with an encoder that encodes images and text

- The result is unstable neural network training producing a model that does not work very well

**AI TOOLS CAN BE USED TO
HELP WRITE SOFTWARE**



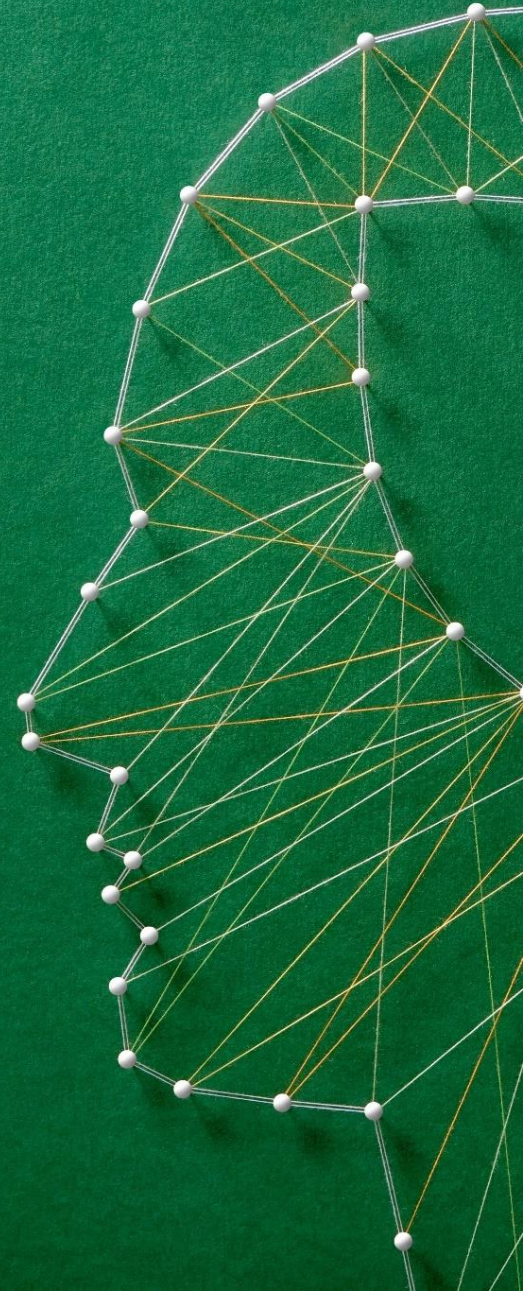
CMS
law·tax·future

Trying to find out why the model failed

The manager asks the team to do testing on the neural network to find out why it is not working very well

Sam says that the objective function needs regularization terms to give stability during training

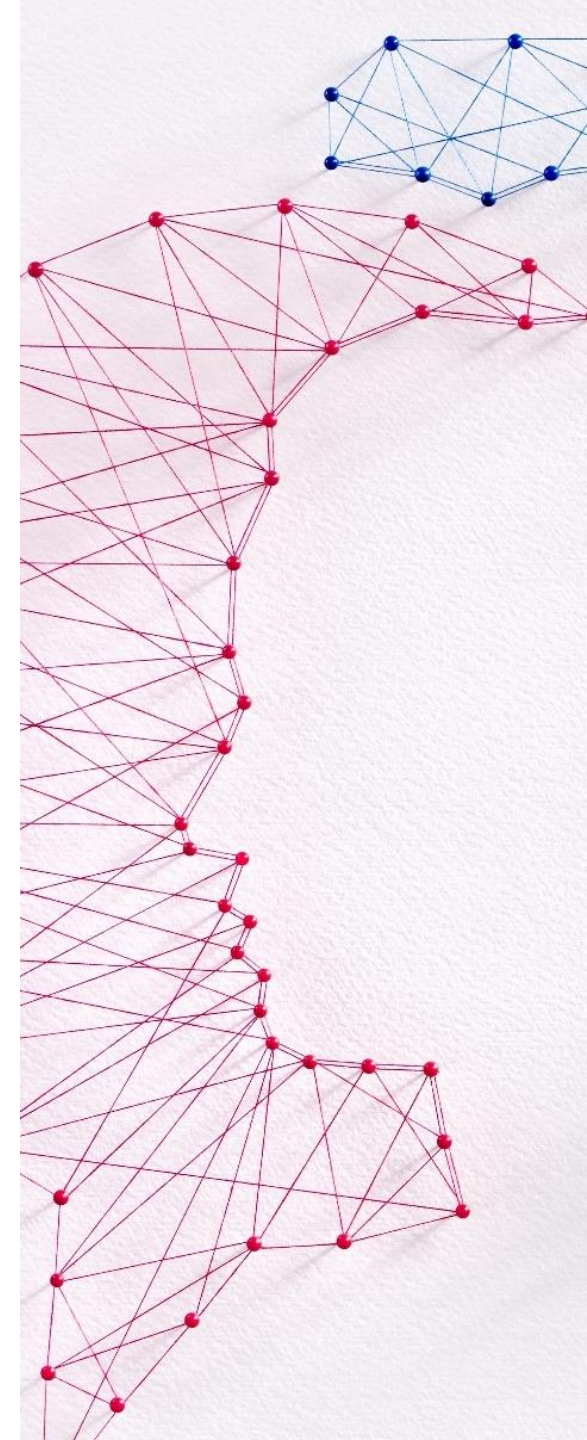
Jamie looks at values of attention vectors in the neural network transformers and discovers some extremely large values. Jamie thinks these are leading to the instability. Jamie finds these happen because of combining the modalities.



Architectural design choices

The whole team tests Jamie and Sam's suggestions. They try different known regularization terms in the objective function but none lead to big improvements.

Alex suggests inserting a normalization layer in particular places in the transformers. It results in good performance. The result is a model that achieves the team's goal.



Result

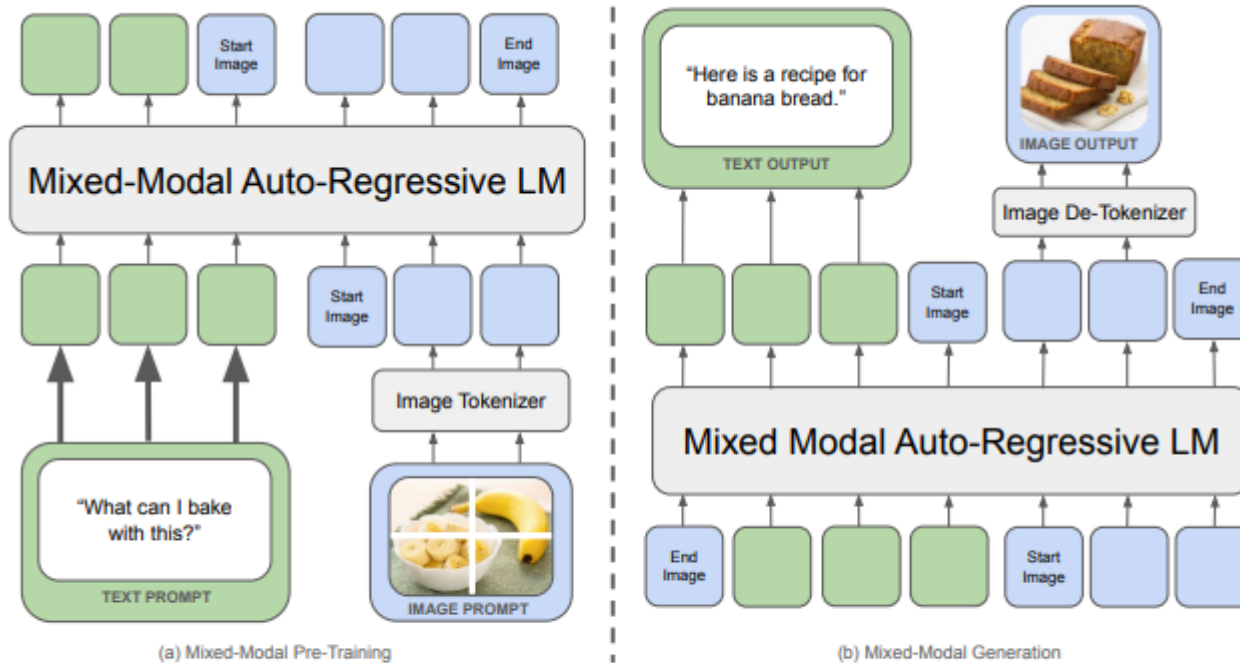
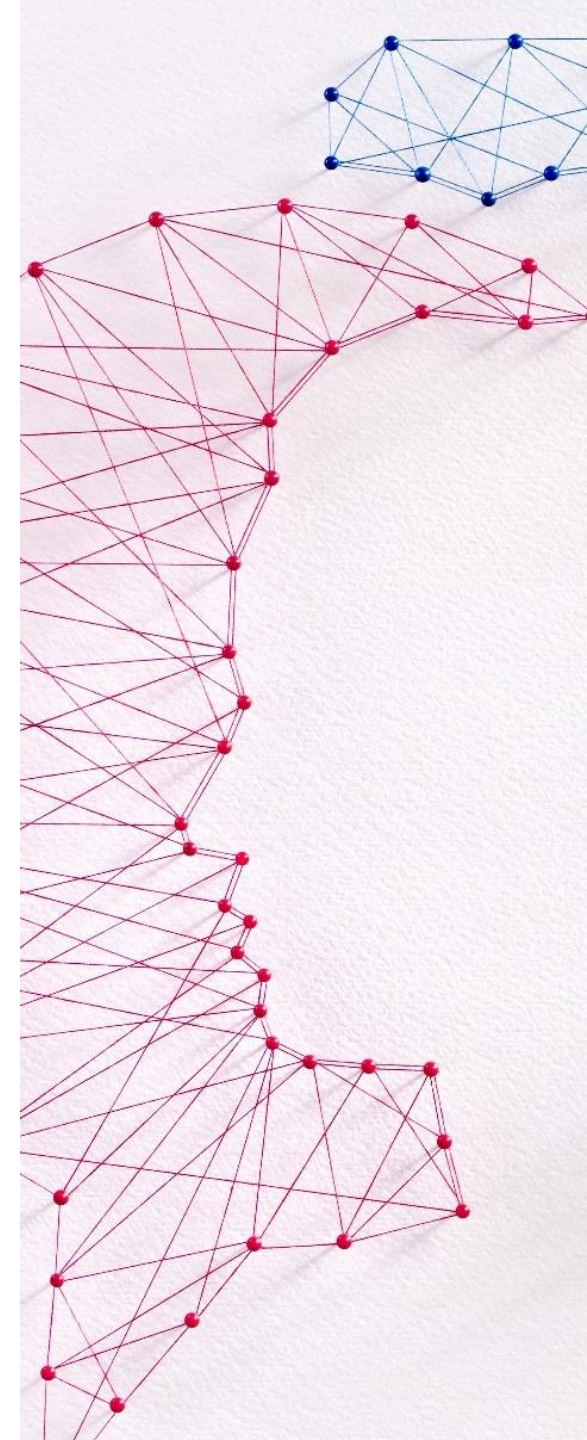


Figure 1 Chameleon represents all modalities — images, text, and code, as discrete tokens and uses a uniform transformer-based architecture that is trained from scratch in an end-to-end fashion on $\sim 10^7$ tokens of interleaved mixed-modal data. As a result, Chameleon can both reason over, as well as generate, arbitrary mixed-modal documents. Text tokens are represented in green and image tokens are represented in blue.

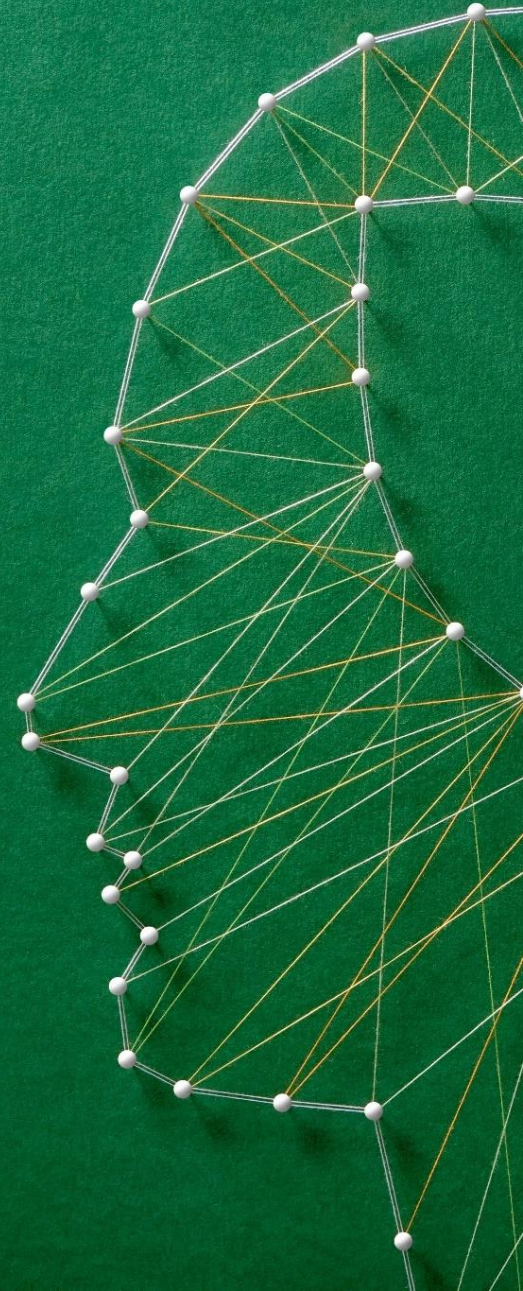
[2405.09818 \(arxiv.org\)](https://arxiv.org/abs/2405.09818)



The manager asks the team to test the model

Sam finds the model works unexpectedly well on medical images

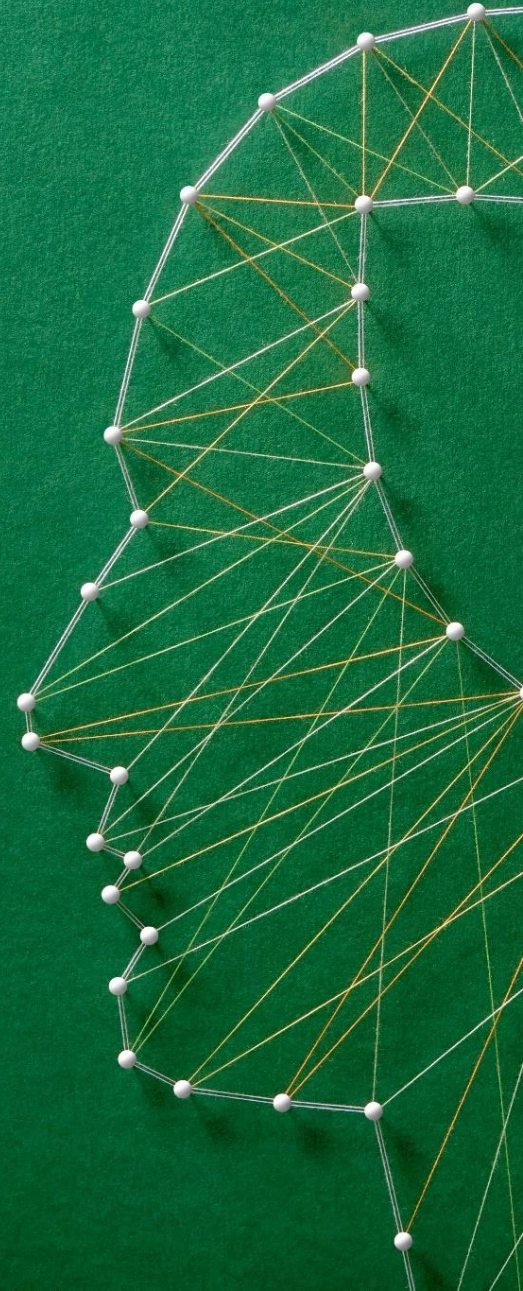
Jamie tests the model on video from the internet and finds reasonable performance.



The patent attorney asks for alternative architectures

The team says there are likely to be alternative architectures that work

One theoretically likely possibility may be to incorporate normalization and softmax into a single layer rather than having them in separate layers

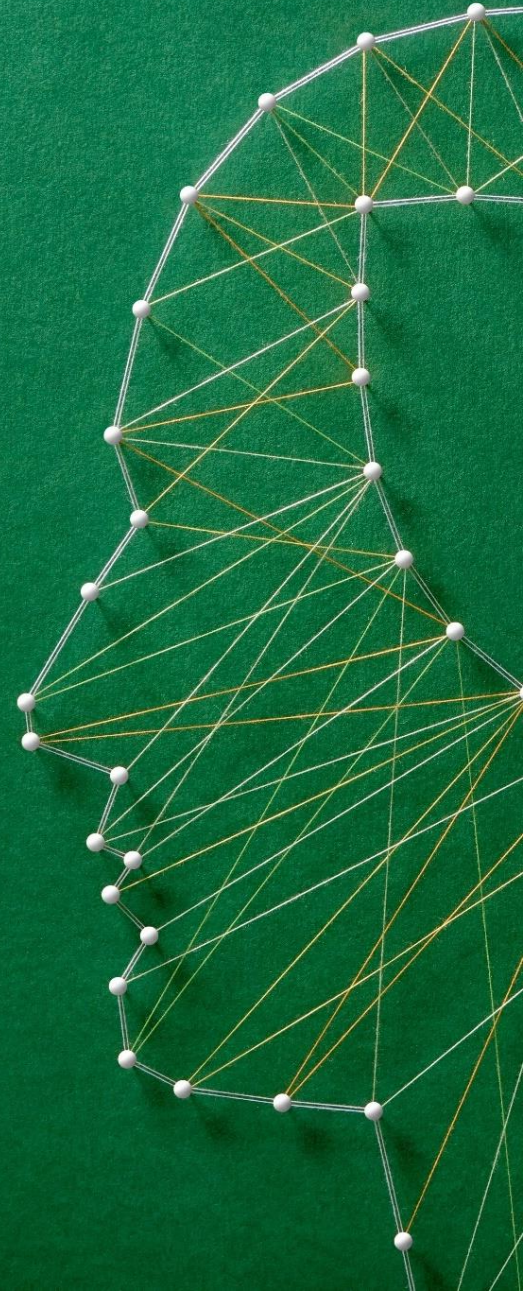


What is the contribution?

Using a single encoder to encode content from different modalities resulting in a multi-modal model that is good at integrating information across modalities and can generate arbitrary sequences of images and text

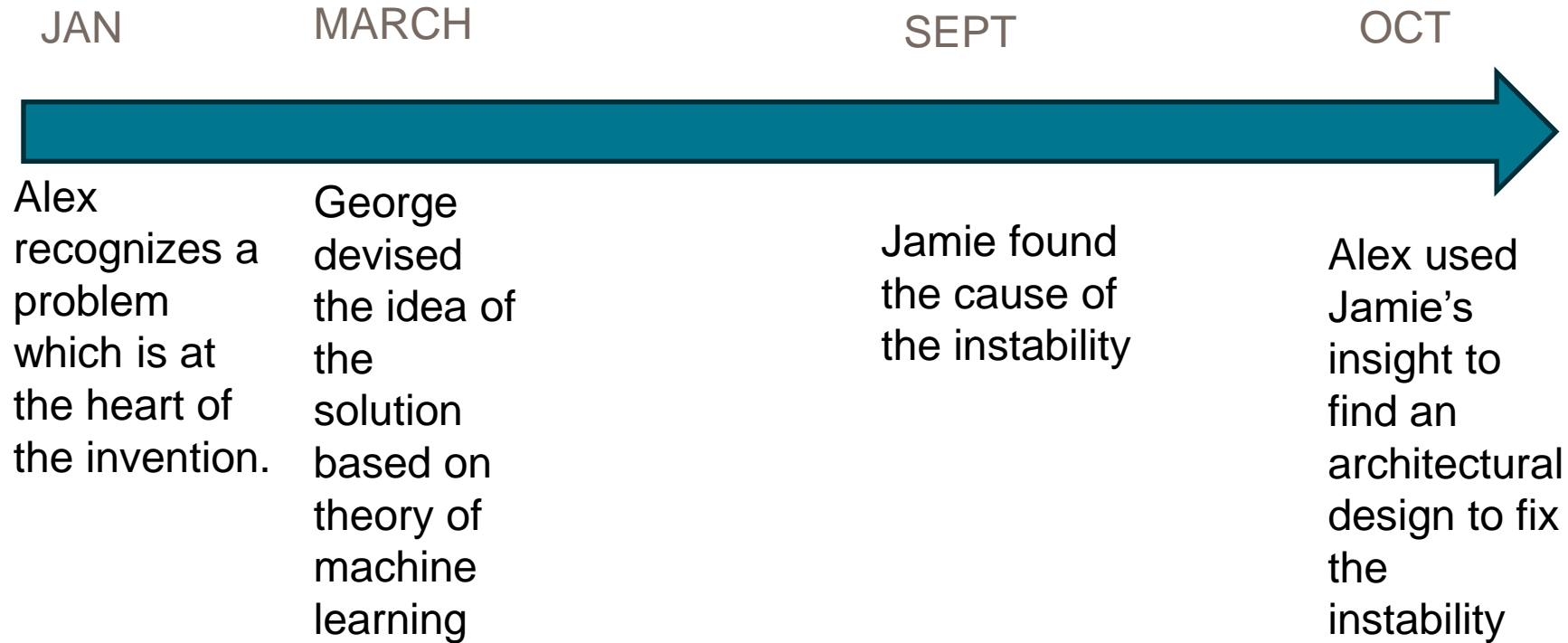
A multi-modal model that is good for generating explanations of medical images

There may be two contributions/inventions

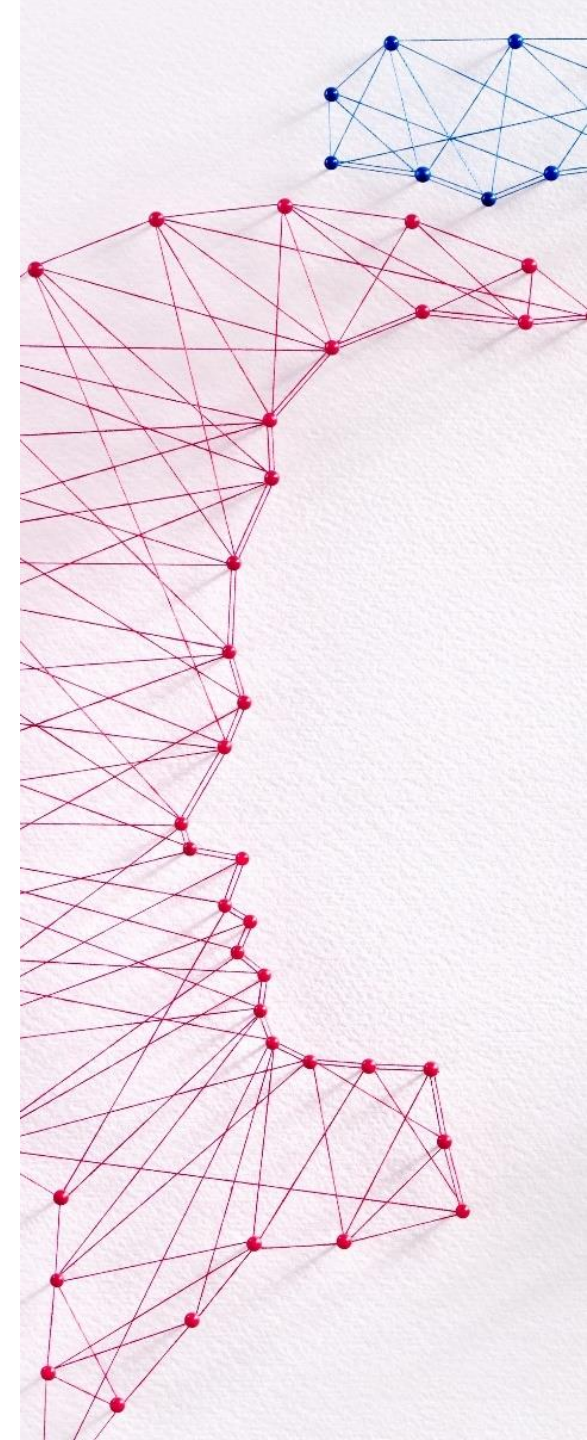


People	Devised the contribution?
The manager assigns task of making a better multi-modal model	Not an inventor. Only has a very general high level aim.
Alex realises that existing multi-modal models are not very good at integrating information across modalities and they can't generate arbitrary sequences of images and text	Is an inventor. Alex recognizes a problem which is at the heart of the invention.
George had an idea to make an encoder that can encode both text and images as theoretically it will enable the model to achieve the team's goal	Is an inventor. George devised the idea of the solution based on theory of machine learning. 7.12.3 UK IPO manual of patent practice
Whole team decided George's idea was the way forward despite the other options	May be inventors as they selected George's idea as a way forward and so perhaps helped devise it?
Jamie found the cause of the instability	Is an inventor. Jamie found a problem which is at the heart of enabling the invention. 7.12.3 UK IPO manual of patent practice
Alex used Jamie's insight to find an architectural design to fix the instability	Is an inventor. Alex found an architectural design that directly solves the problem. The team tried other things that didn't work so it was not straightforward for the Alex to find the enabling architectural design. 7.12.3 UK IPO manual of patent practice
Sam found unexpected performance on medical images	May NOT be inventor. Sam was carrying out specific tests under instruction of another person

Invention timeline



The timeline is long and AI tools are usable along the way



THANKS

for attending this presentation



Patent
Attorneys

Dr Rachel Free

Partner, patent attorney | CMS London

T +44 20 7067 3286

E rachel.free@cms-cmno.com

