

We've made some important changes to our Privacy and Cookies Policy and we want you to know what this means for you and your data.

OK

Find out what's changed

[Home](#)
[News](#)
[Sport](#)
[Reel](#)
[Worklife](#)
[Travel](#)
[Future](#)



[Home](#) | [Video](#) | [World](#) | [UK](#) | [Business](#) | [Tech](#) | [Science](#) | [Stories](#) |

AC



Technology

Google tackles the black box problem with Explainable AI

By Leo Kelion

Technology desk editor

24 November 2019



There is a problem with artificial intelligence.

It can be amazing at churning through gigantic amounts of data to solve challenges that humans struggle with. But understanding how it makes its decisions is often very difficult to do, if not impossible.

That means when an AI model works it is not as easy as it should be to make further refinements, and when it exhibits odd behaviour it can be hard to fix.

But at an event in London this week, Google's cloud computing division pitched a new facility that it hopes will give it the edge on Microsoft and Amazon, which dominate the sector. Its name: **Explainable AI**.

To start with, it will give information about the performance and potential shortcomings of face- and object-detection models. But in time the firm intends to offer a wider set of insights to help make the "thinking" of AI algorithms less mysterious and therefore more trustworthy.

"Google is definitely the underdog behind Amazon Web Services and Microsoft Azure in terms of the cloud platform space, but for AI workloads I wouldn't say that's the case - particularly for retail clients," commented Philip Carter from the consultants IDC.

"There's a bit of an arms race around AI... and in some ways Google could be seen to be ahead of the other players."

MODEL DESCRIPTION



Limitations

The following factors may degrade the model's performance.



Object size: Object size must be at least 1% of the image area to be detected.

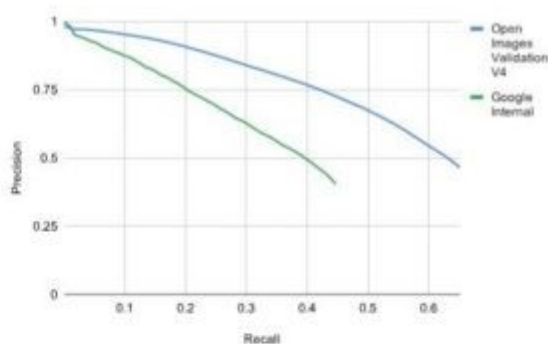


"Things" vs "stuff": Model was designed to detect discrete objects with clearly discernible shapes ("things"), not a group of overlapping objects or backgrounds without clutter ("stuff").



Lighting: Poor or harsh, high-contrast illumination (e.g. nighttime, back-lit, side-lit) may degrade model performance.

PERFORMANCE



Prof Andrew Moore leads Google Cloud's AI division.

He told the BBC the secret behind the breakthrough was "really cool fancy maths".

The transcript below has been edited for clarity and length:

Can you explain what led to Explainable AI?

One of the things which drives us crazy at Google is we often build really accurate machine learning models, but we have to understand why they're doing what they're doing. And in many of the large systems we built for our smartphones or for our search-ranking systems, or question-answering systems, we've internally worked hard to understand what's going on. Now we're releasing many of those tools for the external world to be able to explain the results of machine learning as well. The era of black box machine learning is behind us.

How do you go about doing that - it's not as though you can peer into a neural net and see why an input became an output?

The main question is to do these things called counterfactuals, where the neural network asks itself, for example, 'Suppose I hadn't been able to look at the shirt colour of the person walking into the store, would that have changed my estimate of how quickly they were walking?' By doing many counterfactuals, it gradually builds up a picture of what it is and isn't paying attention to when it's making a prediction.

This is really important, isn't it from a confidence point of view? If we're going to trust not just our businesses, but our lives to artificial intelligence algorithms, it's no good if, when things go wrong, we can't work out why.

Yes. It's really important for societal reasons and fairness reasons and safety reasons. But I will say that no self-respecting AI practitioner would ever release a safety-critical machine learning system without having additional guardrails on it beyond just having Explainable AI.

To be clear, are you saying Google has completely solved the black box problem, or just that you're shining a bit of light in there?

With the new Explainable AI tools we're able to help data scientists do strong diagnoses of what's going on. But we have not got to the point where there's a full explanation of what's happening. For example, many of the questions about whether one thing is causing something or correlated with something - those are closer to philosophical questions than things that we can purely use technology for.

One AI service you aren't offering clients is facial recognition. You've limited yourselves instead to letting clients detect but not recognise faces, with an exception made for those of celebrities. Microsoft and Amazon, by contrast, allow users to build more general facial recognition capabilities into their tools. Why is your approach different?

AMAZON

In general, within Google, we understood how important it is that artificial intelligence is applied responsibly. And so, our chief executive Sundar Pichai commissioned a set of principles that we operate with. They include the fact that we should never be doing harm, and that we should be making sure that the decisions of the systems are unbiased, fair and accountable. As a result of this it does mean that we are very careful. And it does sometimes come across that we are reluctant to just release something and hope that it works because we subject everything to a battery of tests to make sure they are working in a way that's desirable.

Switching tack. Before you took on this role you did AI work for the US Department of Defense. And you joined soon after Google pulled out of a tie-up with the Pentagon to label drone footage - a decision you expressed concerns about. Do you think Google's decision to drop Project Maven was wrong?

That was before my time. So I'm not going to comment on that specific decision.

I will say that one of my roles is to serve on the United States Artificial Intelligence Congressional Commission on AI for national security. And myself, and many other folks throughout the industries understand that we technology providers do have an obligation to help protect countries and societies, as well as producing consumer products as well.

A couple of weeks ago, our chief legal counsel, Kent Walker, **made some comments about Google's great desire** to help out in aspects of national security which will make people safer.

But Google Cloud is ruling out work on weapon systems?

Google's AI principles say that they're not going to be working on offensive weapons systems.

So do you think that Google should be pursuing military or other national security contracts in the future?

I don't want to talk about any specific contracts. But for example, Google is actively helping out with a question of "deepfake" detection, which is this new fear that artificially constructed videos or images might become so realistic that they actually cause societal problems. And so we're partnering with a major government agency in the United States to help deal with that potential.

The decision to abandon Project Maven followed internal opposition to the effort from many Google employees. Do you agree with the view of others, including Microsoft president Brad Smith, that while it's worth listening to workers' concerns you also sometimes need to push back against employee activism?

GETTY IMAGES

One of the things I love about Google, and why I chose to return to Google to work is that it is full of lots of creative voices. And pretty much everything we do, including the design of the shape of buttons on a front-end system, we end up having massive internal arguments about.

Eventually you do have to make a decision one way or the other. The idea of doing top-down management is completely out of Google's culture. But knowing that people are going to disagree and having leadership commit is also something that we are very clear that we do do.

Related Topics

Google

Artificial intelligence

Share this story About sharing

Technology

Elon Musk 'pedo guy' defamation trial to begin

3 December 2019 | **Technology**

TikTok sent user data to China, US lawsuit claims

3 December 2019 | [Business](#)

FBI warns FaceApp poses 'potential threat'

3 December 2019 | [Business](#)

More Videos from the BBC

Recommended by Outbrain

The intimate secret life of nuns

The 'sex tapes' of the most deepfaked woman in the world

The humble rock that changed tech forever

**'I learned I was HIV
positive after my
husband died'**

**Twins 'scarred for life' by
black henna tattoos**

**The moment run-over PC
is pulled to the ground**

Elsewhere on BBC

You Might Also Like

BBC.com

**How much do you really know about
Maldivian culture?**

BBC Culture

**Gentleman's Club: A world that's rarely
seen**

BBC News

**Police chief quit Twitter over hairstyle
abuse**

BBC Future

Why the world is running out of sand

BBC News

Apple worker 'texts himself woman's private photo'

BBC Culture

How vaginas are finally losing their stigma

From Around the Web

Promoted content by Outbrain |

World's Worst Airlines, Ranked

Far and Wide

Backpacks Can't Get More Innovative Than This!

Nordace

Everyone In Netherlands Is Going Crazy Over This Cheap Wifi Booster

techdiscountdeals.com

Affordable SmartWatch Everyone in Netherlands is Talking About

techgadgetdiscounts....

Camera-free smart home alarm

minut.com

Netherlands: People Are Using This Innovative Device For Their Neck Pain

Next Tech

Top Stories

Trump blasts Macron's 'nasty' comments on Nato

Mr Trump accuses the French leader of insulting remarks for suggesting the bloc was "brain dead".

10 minutes ago

Russia to label some journalists 'foreign agents'

1 hour ago

Kim Jong-un opens 'epitome of civilisation' town

3 hours ago

ADVERTISEMENT



INTERNATIONAL ASSOCIATION OF UNIVERSITIES
INTERNATIONAL UNIVERSITY BUSINESS

PRESENTED BY

Degrees of Value

Universities making a difference around the world.

[Watch now](#)

Features



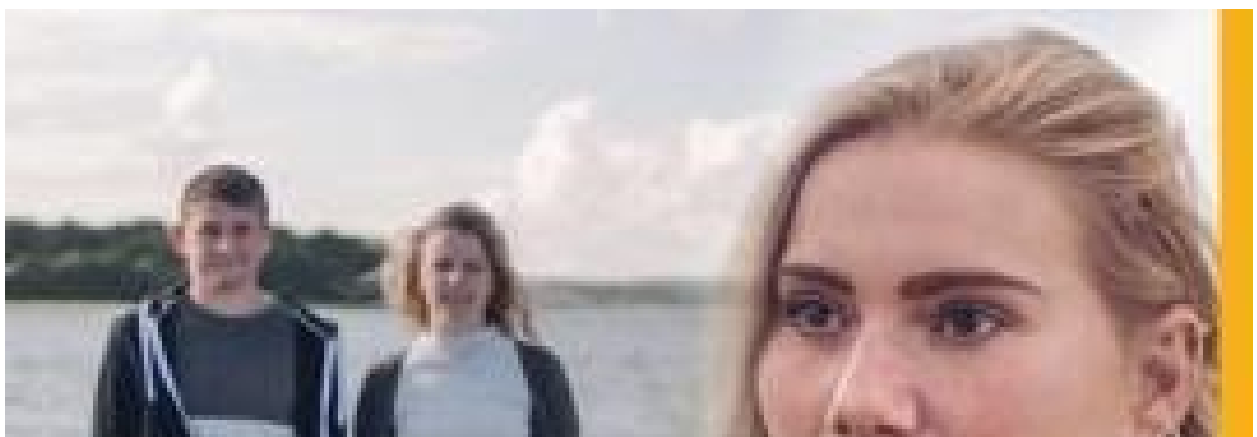
Has a US university cracked student debt?



Italy's Sardines push back against nationalist tide



COP25: Do climate talks actually work?





How seven kids came back from the dead

Why the floods in East Africa are so bad

'Blame the men - not our safe city'

'I was once homeless. Now, I sell homes'

BBC Worklife: One man's 10-year experiment to record every moment

Troubled Nato not in party mood for 70th birthday

Elsewhere on the BBC

Lyrics quiz

Have you been getting these songs wrong?

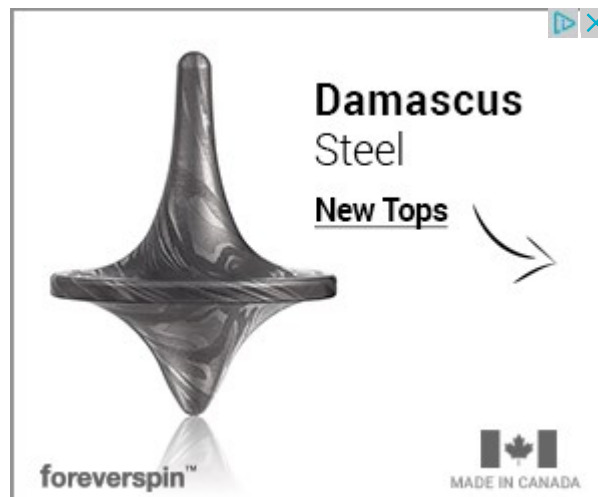
Feeling hot

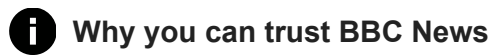
What happens to your body in extreme heat?

Most Read

Donald Trump on UK election: 'I can work with any prime minister'	1
Nato summit: Trump blasts Macron 'brain dead' comments as 'nasty'	2
London Bridge: Jack Merritt was 'phenomenal', says girlfriend in tribute	3
Spotify reveals the decade's most-streamed songs, from Ariana Grande to Drake	4
Loughton hit-and-run: Murder arrest over fatal crash	5
Ellie Goulding: 'I used alcohol to be more funny and interesting'	6
Elon Musk 'pedo guy' defamation trial to begin	7
Mental As Anything singer Andrew 'Greedy' Smith dies aged 63	8
North Korea unveils town hailed as 'epitome of civilisation'	9
Dog starts house fire in Essex by turning on microwave	10

ADVERTISEMENT





BBC News Services

On your mobile

On your connected tv

Get news alerts

Contact BBC News

Explore the BBC

Home

Sport

Worklife

Future

Music

Weather

News

Reel

Travel

Culture

TV

Sounds

Terms of Use

Privacy Policy

Accessibility Help

Contact the BBC

Advertise with us

About the BBC

Cookies

Parental Guidance

Get Personalised Newsletters

Ad choices

Copyright © 2019 BBC. The BBC is not responsible for the content of external sites. **Read about our approach to external linking.**