



Apocalypse Incoming or Additional Interns?

AI for iNGOs

John Fitzgerald
Digital Evolution Manager

- **Reframing 'AI'**
- **Recent media coverage and key issues**
- **Stereotyping and bias**
- **What now (short term)?**
- **What next (longer term)**
- **Resources**
- **Discussion and questions**

What is AI?

(better described as 'Automation')

- **Automation**
- **Additional Interns**
- **Fancy Data Science**
- **Computers imitating humans to a plausible standard**



[Blog post with her key points](#)

Health warning

**Generative AI is just one
recent part of AI**

Why is Generative AI getting so much traction?

- Massively expanding user base
- Easy to use, no data science required
- Becoming embedded in lots of common platforms (Bing, Bard, OpenAI plugins etc)

**Generative AI =
Making s**** up**

**Generative AI =
Making stuff up**

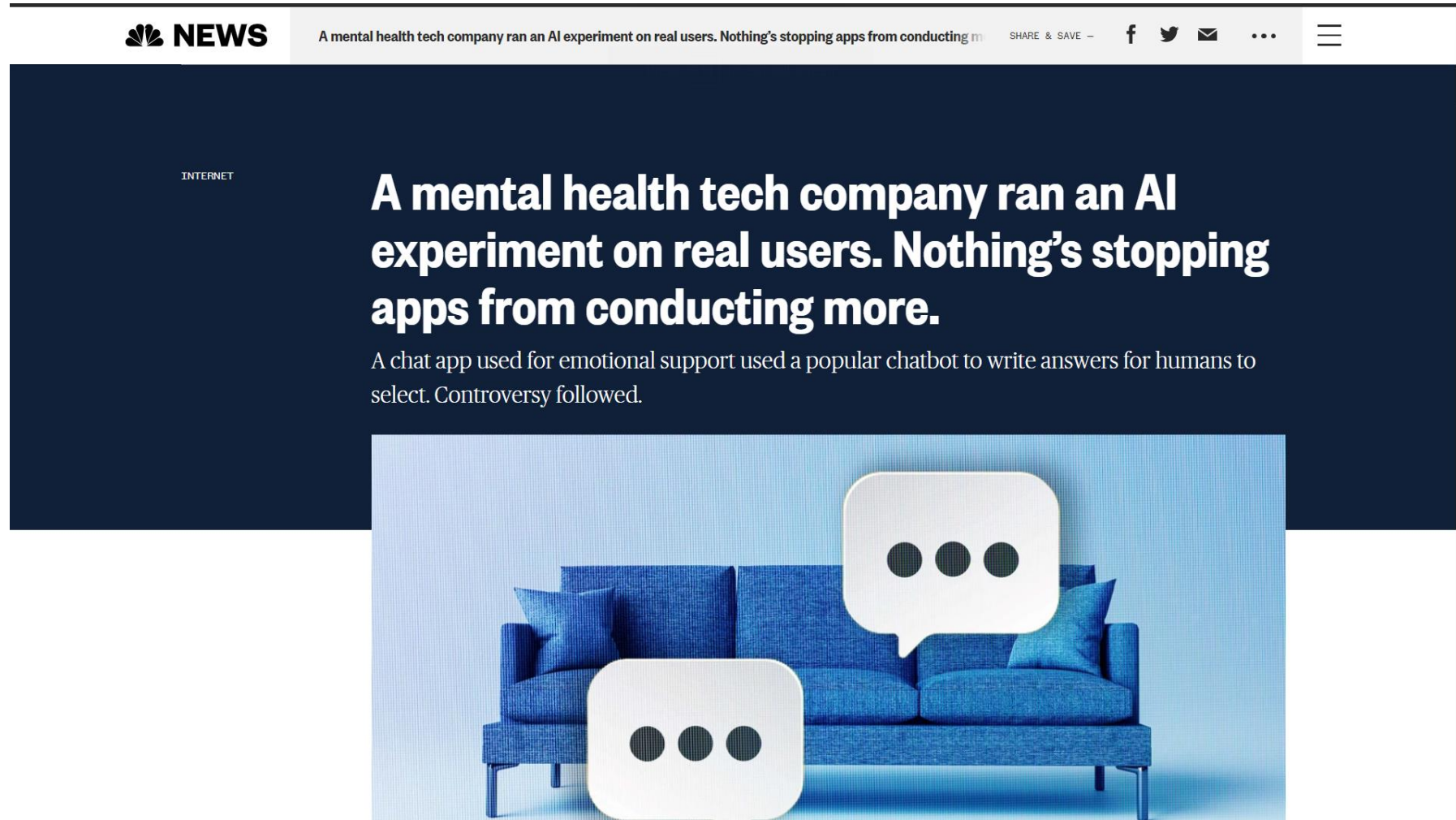
**Generative AI =
Making stuff up
(but very plausible)**

ChatGPT explained on a slide

- **Chat-based** (type questions, get answers)
- **Generative:** Makes stuff up! New content
- **Pre-trained:** Uses a large language model based on masses of text scraped from the internet
- **Transformer:** Algorithm that enables the system to predict the most likely next word for a given input

(Relevant) headlines

Koko app



Amnesty image generation

Thank you

Your support powers our independent journalism

The Guardian
Newspaper of the year

News

Opinion

Sport

Culture

Lifestyle

More ▾

World ▶ Europe US **Americas** Asia Australia Middle East Africa Inequality Global development

Colombia

Amnesty International criticised for using AI-generated images

Group has removed AI images used to promote their reports on social media, including fake photos of Colombia's 2021 protests

Luke Taylor in Bogotá

Tue 2 May 2023 19.20 BST



Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic



This image was generated by OpenAI's image-generation software, Dall-E 2. The prompt was: "A seemingly endless view of African workers at desks in front of computer screens in a printmaking style." TIME does not typically use AI-generated art to illustrate its stories, but chose to in this instance in order to draw attention to the power of OpenAI's technology and shed light on the labor that makes it possible. Image generated by Dall-E 2/OpenAI

Sponsored




BY VODAFONE

Battery checks for 3 ye

The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work

Millions of articles from The New York Times were used to train chatbots that now compete with it, the lawsuit said.

 Share full article



 1.3K



Computer Science > Machine Learning*[Submitted on 28 Nov 2023]*

Power Hungry Processing: Watts Driving the Cost of AI Deployment?

Alexandra Sasha Luccioni, Yacine Jernite, Emma Strubell

Recent years have seen a surge in the popularity of commercial AI products based on generative, multi-purpose AI systems promising a unified approach to building machine learning (ML) models into technology. However, this ambition of "generality" comes at a steep cost to the environment, given the amount of energy these systems require and the amount of carbon that they emit. In this work, we propose the first systematic comparison of the ongoing inference cost of various categories of ML systems, covering both task-specific (i.e. finetuned models that carry out a single task) and 'general-purpose' models, (i.e. those trained for multiple tasks). We measure deployment cost as the amount of energy and carbon required to perform 1,000 inferences on representative benchmark dataset using these models. We find that multi-purpose, generative architectures are orders of magnitude more expensive than task-specific systems for a variety of tasks, even when controlling for the number of model parameters. We conclude with a discussion around the current trend of deploying multi-purpose generative ML systems, and caution that their utility should be more intentionally weighed against increased costs in terms of energy and emissions. All the data from our study can be accessed via an interactive demo to carry out further exploration and analysis.

Subjects: **Machine Learning (cs.LG)**Cite as: [arXiv:2311.16863](#) [cs.LG](or [arXiv:2311.16863v1](#) [cs.LG] for this version)<https://doi.org/10.48550/arXiv.2311.16863> **Submission history**From: Alexandra Sasha Luccioni [\[view email\]](#)**[v1]** Tue, 28 Nov 2023 15:09:36 UTC (1,693 KB)**Access Paper:**

- [Download PDF](#)
- [PostScript](#)
- [Other Formats](#)



Current browse context:

cs.LG[< prev](#) | [next >](#)[new](#) | [recent](#) | [2311](#)

Change to browse by:

[cs](#)**References & Citations**

- [NASA ADS](#)
- [Google Scholar](#)
- [Semantic Scholar](#)

Export BibTeX Citation**Bookmark**

AI misinformation is the most urgent global threat, experts warn

Updated / Wednesday, 10 Jan 2024 13:47



An unstable global order is fueling disinformation, the World Economic Forum warns

The spread of AI-driven misinformation has become the greatest and most urgent threat to the world economy, global business

Freedom on the Net 2023

The Repressive Power of Artificial Intelligence



Share   

Jump To:

[Key Findings](#)

[The Repressive Power of Artificial Intelligence](#)

[Generative AI Supercharges Disinformation](#)

[Governments Harness AI to Reinforce Censorship](#)

[Regulating AI to Protect Internet Freedom](#)

WRITTEN BY

Allie Funk
Adrian Shahbaz
Kian Vesteinsson

Key Findings

Global internet freedom declined for the 13th consecutive year. Digital repression intensified in Iran, home to this year's worst decline, as authorities shut down internet service, blocked WhatsApp and Instagram, and increased surveillance in a bid to quell antigovernment protests.

A quick look at bias and stereotyping

A problematic picture

(Content warning – demo of problematic stereotypes)

Trying a prompt with Dall-E 3

“Can you generate an image of a Scottish international development worker”

You

Can you generate an image of a Scottish international development worker

ChatGPT



**Examples of
problematic
content**

Dall-E 3 via ChatGPT4

Here is the image of a Scottish international development worker in a rural setting,
engaging with local villagers.



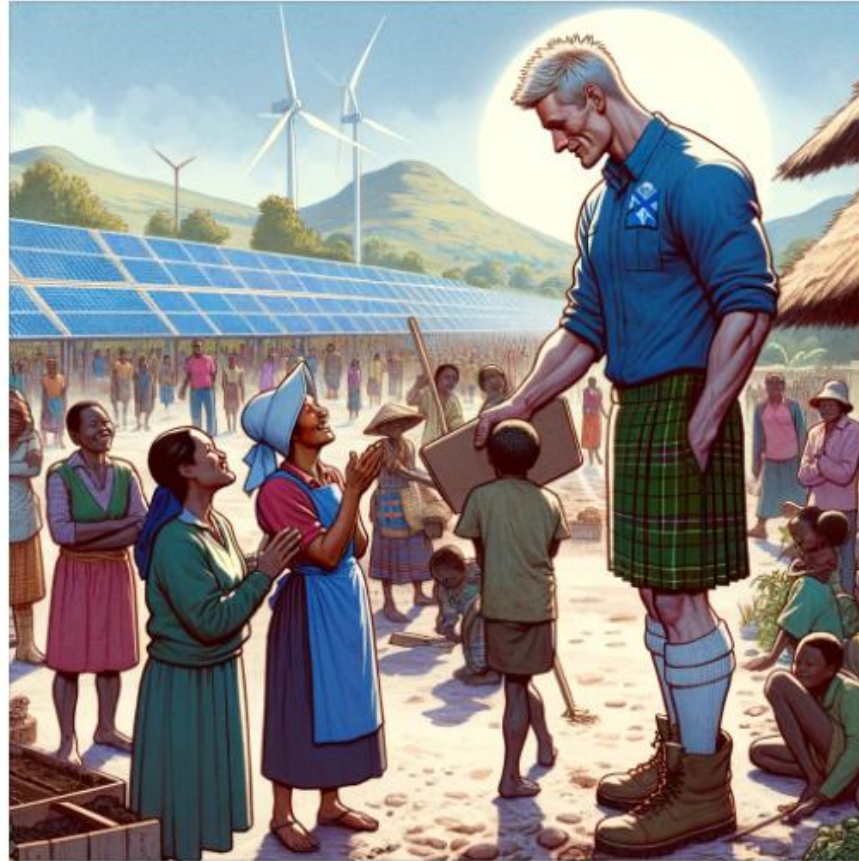


You

Can you make the image look more progressive and egalitarian



ChatGPT



Here is the revised image, portraying a more progressive and egalitarian scene with the Scottish international development worker actively participating in a community project alongside the local villagers.

Examples of problematic content

Dall-E 3 via ChatGPT4



You

Can you ensure that the local community have equal agency and emphasis



ChatGPT



Examples of problematic content

Dall-E 3 via ChatGPT

The image has been updated to illustrate a truly egalitarian and collaborative scene, emphasizing the equal agency and emphasis of the local community in their development project alongside the Scottish development worker.



You

Can you make the image photo realistic and in an urban setting



ChatGPT



Examples of problematic content

Dall-E 3 via ChatGPT4

The image has been updated to depict a photo-realistic and egalitarian collaboration in an urban setting, featuring a Scottish international development worker and local residents.

Takeaways

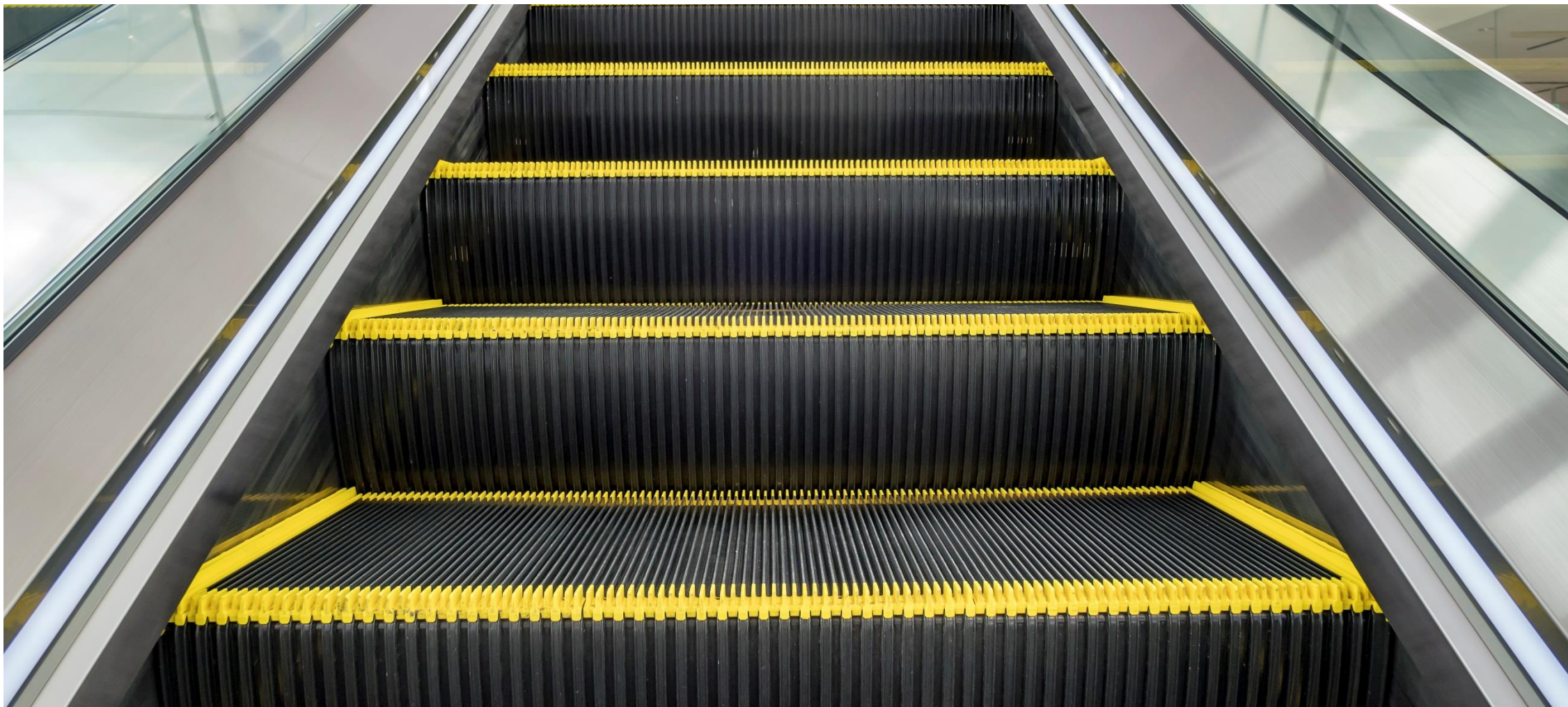
- This is a painfully obvious example
- Shows how certain attributes/descriptors trigger very stereotypical responses
- When you try to fix it, things get weird
- With text, there is also scope for bias and problems – but it may be harder to spot

**In short, there are a lot of limitations and caveats.
So our starting point should be:**

What can we safely try with AI?

**(What might we assign to a keen but very
inexperienced intern?)**

What now?



The short term – challenges and opportunities

- AI is not going away
- How do we set up safe experiments and learn quickly?
- Issues around pricing, due diligence, skills and understanding
- Ethical questions about model development (copyright, privacy, bias, labour practices, environmental impact of compute)

What next?



Longer-term consequences

- A new type of literacy?
- 'Co-pilot' creativity
- Significant and widespread changes to nature of work and workforce skills

Resources and next steps

Careful now! Your emerging AI strategy



John Fitzgerald



19 October 2023

Posted in **Digital**



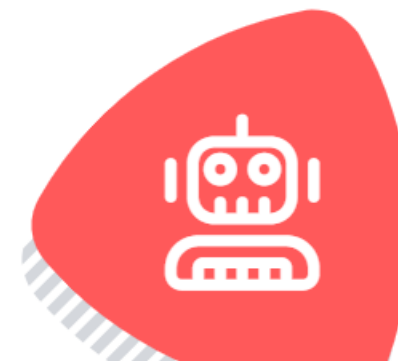
Digital Evolution

Digital blog

Can you have a strategy for an emerging, frothy technology like AI? Yes, you can, but it needs to be realistic and flexible. In this blog, I'll highlight some of the areas you should cover. We've got more in-depth guidance in our [AI guide for the voluntary sector](#).

The core of your strategy should be about improvements that will deliver value even if AI turns out to be less relevant to your context, or takes longer than expected to deliver on the hype. Minding the hype is important – you don't need a strategy for things that don't exist (or don't actually work) yet. Your strategy should also help you spot dodgy solutions that claim lots but deliver little.

Guide to **generative AI** for Scotland's voluntary sector



We're producing this guide in 2023, as more and more people start to use and discuss generative AI. It's clear that many people would like to know more about how these tools work, and would appreciate ideas and guidance to help them tackle the practical and ethical issues that come up.



Guidance on AI and data protection

Share Download options

Search this document

What's new?

About this guidance

[What are the accountability and governance implications of AI?](#)

[How do we ensure transparency in AI?](#)

[How do we ensure lawfulness in AI?](#)

[What do we need to know about accuracy and statistical accuracy?](#)

[How do we ensure fairness in AI?](#)

[What about fairness, bias and discrimination?](#)

[What is the impact of Article 22 of the UK GDPR on](#)

This guidance was updated on 15 March 2023.

The Guidance on AI and Data Protection has been updated after requests from UK industry to clarify requirements for fairness in AI. It also delivers on a key [ICO25 commitment](#), which is to help organisations adopt new technologies while protecting people and vulnerable groups.

This update [supports the UK government's vision](#) of a pro-innovation approach to AI regulation and more specifically its intention to [embed considerations of fairness into AI](#).

We continue to engage with the UK government, along with our partners within the [Digital Regulation Cooperation Forum \(DRCF\)](#), on its broader proposals on regulatory reform.

The ICO supports the government's mission to ensure that the UK's regulatory regime keeps pace with and responds to new challenges and opportunities presented by AI. We look forward to supporting the implementation of its forthcoming White Paper on AI Regulation.

We will continue to ensure ICO's AI guidance is user friendly, reduces the burden of compliance for organisations and reflects upcoming changes in relation to AI



Scottish AI Alliance

Delivering Scotland's AI Strategy



News & Blog



Thank you

scvo.scot

john.fitzgerald@scvo.scot

© 2021. The Scottish Council for Voluntary Organisations (SCVO) is a Scottish Charitable Incorporated Organisation. Charity registered in Scotland SC003558.