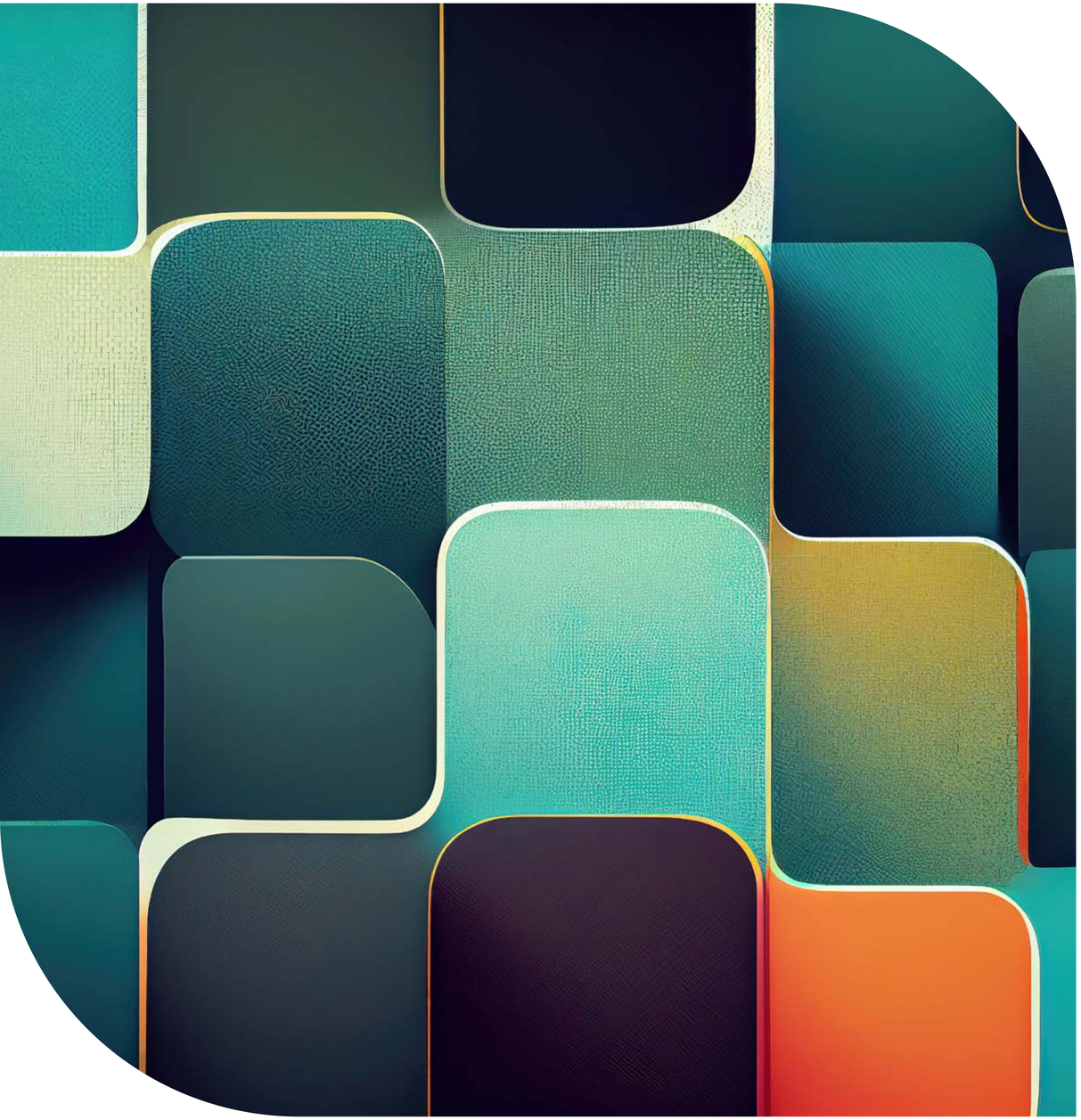
AI & RESPONSIBLE INNOVATION: WHAT'S NEXT?



CONTENTS

FOREWORD

Dr Iain Brown
Head of Data Science,
SAS UK & Ireland

It's now 20 years since the film, The Minority Report, was released – and during that time, we've seen rapid development in artificial intelligence (AI), machine learning (ML) and other data-driven technologies.

While the film might be based on the predictions of psychics, rather than the AI we see today, it raises important ethical questions about how predictive technology could be used, both now and in the future, and what safeguards are needed.

Most people recognise there are risks if AI is unchecked, and could lead to situations where decisions fail to respect privacy, are shrouded in secrecy, and unfairly disadvantage groups or individuals. In the wrong hands, AI could also be used to proliferate fake news, and disrupt entire economies and societies, as data scientist and thought leader Kirk Borne points out on p4 of this report.

But while some of the predictions made 20 years ago have come to bear (such as **personalisation in retail**), we've also seen a healthy amount of push back, which is one of the reasons not all of the technology considered 'future state' has been adopted.

Legal uncertainty has made organisations cautious too, so we're not yet seeing the full power of AI. Nobody wants to be the first to fail because of the reputational damage it entails.

Upcoming legislation, combined with sharing of best practice, is expected to bring some clarity. Kalliopi Spyridaki, Chief Privacy Strategist, SAS EMEA & Asia Pacific, looks at the legislation set to govern AI on p10. Meanwhile, Prathiba Krishna, a data scientist at SAS UK & Ireland, suggests that the right safeguards promote 'responsible innovation' and build trust among consumers on p6.

When using AI, the question to ask is, 'how does it benefit people?'.

There may be obvious advantages, such as being able to accurately predict the likelihood of someone suffering from a serious disease, as SAS UK & Ireland data scientist Niall Larkin discusses on p8. But he also points to other benefits such as removing repetitive and labour-intensive tasks, so employees can focus on more strategic and innovative ones.

There has been tremendous interest recently in Open AI's ChatGPT, the chatbot which uses natural language processing to generate written text that most people would think came from a human. Niall also touches on this example.

Greater transparency around AI models will help to ensure that people can make informed decisions about how their data is used – and they'll quickly withdraw their consent if they don't see a benefit, or the organisation breaches their trust.

In another 20 years, we can expect to see some rules governing AI become better-defined. In addition, organisations need to be proactive in preparing their frameworks and technology so they can take advantage of the advances in AI. So films like The Minority Report might feel more like fiction than reality.

At SAS, we align ourselves to six core principles for responsible innovation, starting with human-centricity – i.e. promoting human well-being, agency and equity. This is followed by:

- **Inclusivity:** ensuring accessibility and including diverse perspectives and experiences
- **Transparency:** explaining and instructing on usage openly
- **Accountability:** proactively identifying and mitigating adverse impacts
- **Robustness:** operating reliably and safely
- **Privacy and security:** respecting the privacy of all data subjects.

These principles guide our AI development, with strict model governance in place to ensure fair and equitable decisions. We continually test our models against challenger models, and optimise them as new data becomes available to ensure the best outcomes for everyone.

NAVIGATING THE GRAND AI EXPERIMENT

Dr. Kirk Borne, data scientist, global speaker and founder of the Data Leadership Group, looks at how AI is developing – and what safeguards are needed to ensure it's used for good.

The Wild West is a good analogy for where AI is right now. It's become so powerful, and so pervasive, that it's increasingly difficult to tell what's real or not, and what's good or bad.

High-profile examples, like the AI-generated image of Chris Farley, might be a talking point – but we can imagine what damage could be caused if the technology gets into the hands of malicious agents, operating in, or for, hostile regimes.

Of course, it's not only the outright criminals we need to be wary of. With its smart decision-making capabilities, AI is being adopted by businesses and governments faster than it can be regulated. The question is, how can they avoid the unintended biases that can creep into models and unfairly disadvantage certain groups?

While sensitive characteristics, like age, gender and race, are relatively easy to engineer out of models, there are other seemingly-innocuous features that might result in biases.

Knowing what kind of operating system someone is using, for example, is essential for website functionality, yet it can also lead to unfair assumptions about them. An iOS user might be deemed more affluent than an Android user and therefore targeted with higher-priced products.

As I explain later, the widespread use of AI is effectively a grand experiment on human subjects – and as such, we all need to be confident that it is ethical, responsible and explainable.

Current landscape

At the moment, AI is faithfully following Gartner's technology hype cycle, from the peak of inflated expectations to the trough of despair, and now we're heading towards the plateau of productivity.

The algorithms have become extremely powerful, with deep learning and neural networks vastly improving our image recognition and natural language processing capabilities.

We're now moving past the research-intensity phase to the application of AI and how it can create value in the real world. This is where you see the productivity – but you can also run into trouble if it isn't implemented ethically.

It's important to remember that AI can be applied, uncontroversially, in multiple ways. In cold chain logistics, where products need to be maintained at the correct temperature, the environment is monitored by sensors and the data used to build predictive models for demand, and arrival times.

Few people are concerned about data relating to productivity, revenue or weather, all of which is important in business and policy-making. Understandably though, many are worried about how their personal data is used and the life-changing decisioning capability of AI to determine, say, whether they'll be successful in their mortgage applications or receive the correct medical diagnosis.

Keeping an eye on the AI

AI is advancing so rapidly that it's difficult for the regulators to keep pace. The EU offers a **guide to ethical AI** but companies aren't waiting for heavy-handed regulation – they're self-policing because they know their success and reputation depends on doing the right thing.

One financial institution I know is trying to fix its credit scoring models by intentionally removing any variable that could lead to someone scoring lower because of their race or gender.

Of course, AI is just mathematics, which is neither ethical nor unethical. If two things co-occur then there's a higher probability that other things will co-occur. What this company has done is reverse engineer its models by taking the results and seeing if they can infer the missing features. Is there something embedded in the solution that shouldn't be there?

This is why, instead of relying on automation to validate AI models, we are seeing more humans in the loop because only they can ensure the AI is transparent and explainable.

I'd advocate the use of what I call AI-ops, similar to **machine learning ops**, where the performance of the model is continually monitored and measured against KPIs such as 'no more than 5% of the model should be informed by what you can't know'.

Genie out the bottle

As with any new technology, we need to push the boundaries of what's possible to see whether an application is ethical or not. OpenAI's GPT-3 uses **deep learning to generate human-like text** – and it's telling that the source code hasn't been released. In the wrong hands, it could lead to anything from students cheating on homework tasks, to a proliferation of fake news – such as false financial information or medical research – aimed at discrediting people and disrupting society.

This is why people in the know, the data experts, are getting worried. People talk about the **Turing Test**, but what worries me more is a model that deliberately fails the Turing Test, that is, doesn't let on that it knows more than you. A tool like **Google Duplex** could feel scary if people don't know AI is being used – yet, with transparency, it is also a highly-convenient way of managing daily tasks.

The rights of human subjects

Having taught data ethics in universities, I believe that we can learn a lot from the medical research community, specifically, how clinical trials are conducted.

The Tuskegee Syphilis Study and experiments conducted by the Nazis during the Holocaust are among the most horrifying chapters in our recent history, involving human subjects who didn't provide consent and who were physically and mentally harmed in the process.

The ethical frameworks that have been developed since, including the Nuremberg Code and the principles from the Belmont Report, help to safeguard participants and prevent future atrocities.

> It struck me that AI is a grand experiment on humanity – and we're all the ultimate human subject.

We therefore need to incorporate the core principles of human subject research, including:

1. **Autonomy and informed consent**
2. **Do no harm**
3. **Fairness and equity (i.e., the benefits and risks are shared across the participants).**

Left to its own devices, one study suggests that **AI 'can't be trusted'** because it's unable to ensure 'fairness, privacy and robustness' by itself. This brings us back to keeping humans in the loop at all times because only we can make a judgement about something that affects real people. To be ethical, AI must be explainable, so people can make an informed decision with clarity, depth and understanding.

RESPONSIBLE INNOVATION

Prathiba Krishna, a data scientist at SAS UK & Ireland, explains why fairness and equity must be built into every AI model from the start – and why ethics fuels rather than stifles innovation.

Everyone has the right to question, challenge or raise concerns about the technologies rapidly becoming part of our lives, especially ones as powerful as AI.

Bias in AI algorithms is one of the biggest challenges we still need to overcome – but there are further complexities. Do people feel deceived if, for example, they feel an emotional connection with a chatbot that appears to display empathy? Or should they be suspicious when a machine makes a decision on their loan or mortgage application – or a medical diagnosis – with limited human intervention?

As we adapt to the smart world, we all have our own understanding and tolerances around how our data and AI is used.

Younger generations, who've grown up with technology, generally understand how their data informs, say, targeted advertising and are more comfortable with it. Older generations, however, may want more reassurances that any reputable organisation will aggregate and anonymise their personal information, so they can't be identified as individuals.

> To win consumers' trust, a clear and strictly-governed framework that promotes privacy, respect, transparency and consent is needed, with space for rapid yet controlled experimentation. Because it is only by testing out new models and algorithms that the most effective and ethical ones can be developed.

How and why

As data scientists, we should continually question the quality and balance of the data, and be open to how this changes as more data becomes available. An ethical model must be both interpretable (understanding how a model or algorithm came to its conclusion) and explainable (understanding why it did).

We have to be storytellers who're able to clearly articulate what's happening. Whereas a data engineer is comfortable with more technical detail, another stakeholder, such as a compliance manager, needs decisions to be explained in terms they can understand, so if something goes wrong, they can work backwards and see why. With new regulations from the **UK government** and the **EU** on the horizon, companies will soon be required to explain to consumers, in plain English, how decisions are made.

Over the past few years, industry practices have changed and there are now more ways to visualise, assess and detect potential biases throughout the model lifecycle.

As data science practices have matured, we're able to look for potential bias at a much earlier stage and therefore reduce the operational, financial and emotional risk. Whereas previously we'd look for biases after the model had been built, and the data had been pulled to a separate graph, we now have in-built technologies and attachments capable of doing this from the start.

In the coming years, I have no doubt there will be many more ways to assess the AI and more variables to consider.

Ethical AI runs end-to-end, ensuring accountability and auditability at every stage – and the right safeguards drive **responsible innovation**, so more people ultimately benefit from it.

Guided by the SAS Data Ethics Practice, we apply this principle to make sure models are fair, equitable and human-centric, from development to deployment. Our goal is to empower people through the data-driven technologies that can enhance many areas of our lives.

I firmly believe that controlled risk is a good thing. Data scientists need an R&D unit to test the tools and the freedom to fail fast to find out what does and doesn't work before the technology is rolled out to organisations and consumers. If you don't experiment, you won't find the optimal model. Once deployed, models should be continually monitored, and challenger models run in parallel, to determine which is the most effective and the most ethical. In some cases, the challenger model will become superior and take over.

Making ethics a priority

Almost every industry today, from the highly-regulated banking, insurance and healthcare sectors, to recruitment, retail and many more, are impacted by AI ethics.

Most organisations are talking about ethics although they're not always sure what it means for them in practice. Often, it isn't top of their list because they also have data infrastructure, production and IT challenges to deal with – but this is where we bang the drum and say that it must be a priority because sooner or later, practices will be regulated.

A partner like SAS has an important role to play in educating these organisations about AI ethics, and providing the tools and guidance for them to innovate responsibly. For example, I have demos that I can run at any time to show what bias could look like.

At the start of every new data science project, we have a conversation about ethics to find out their views, brainstorm ideas and understand their business rules. Whereas some industries might remove sensitive characteristics like age and gender from their models, medical industries have to include them to accurately predict the likelihood of disease. The question we need to understand is how do their business rules create bias and how could they be mitigated?

Fundamentally, consumers must agree to their data being captured and used in a certain way; to know that they're talking to a chatbot not a human; and that their personal details are protected. With an ethical framework in place, which incorporates privacy, respect, transparency and consent, companies can innovate both quickly and responsibly.

GROUNDING AI IN REALITY

Niall Larkin, Data Scientist at SAS UK & Ireland, explores where AI is going – and why the future might not be as dystopian as some people think.

AI, including natural language processing (NLP) has become so sophisticated that it's difficult not to ascribe human qualities to it, or even believe that it's sentient. Google engineer Blake Lemoine may have been dismissed for his claims that one of the company's chatbots had expressed human emotions – yet it's easy to see how people could imagine it soon becoming a reality.

There is no doubt that some of the latest innovations in AI are ground-breaking but we shouldn't conflate complexity with sentience. What we're seeing is multiplication of matrices that certainly give the illusion that humans could be replaced with AI, but the reality is a long way from science fiction.

Believe the hype?

There's a lot of hype around AI, for example with art applications like DALL-E and Midjourney. That hype though tends to fade quite quickly, as people move onto the next big thing.

Images created using AI can be extremely impressive but the idea that the technology removes the need for artists doesn't sit well with many people.

It's true that AI, by itself, may never be able to encapsulate the ideas and emotions of artists over history – but it can still be a useful creative tool for artists. I've used Midjourney, and you can create something fantastic with it. Where I see these tools adding value is in sectors like marketing and graphic design. Again, it won't replace the creativity that humans bring but it could help us to brainstorm and create quickly highly-useful new concepts for logos or products.

One AI application that has generated a huge amount of debate is ChatGPT, which provides convincingly human responses to almost any prompt. One reason for the hype around it is the unprecedented ease of access – it can be put in the hands of anyone who wants to create a free user account on the service. Among the benefits are that the threshold for organisations to realise the value of the technology is lowered, and new innovative business models can arise with a focus on strengthened customer and citizen relations. Yet with such ease of access, concerns have been raised about it being used inappropriately – e.g. by students using it to complete assignments.

ChatGPT is a form of generative AI, that is AI which produces or generates text, images, music, speech, code or video. It uses natural language processing to create text that feels almost like it's come from a human rather than a typical chatbot. It was trained on a wide variety of published materials, including books (fiction and non-fiction), web pages, social media, and scientific journals. The model incorporates an astronomical number of pathways that allow it to predict a good answer from almost any prompt. It was created by Open AI, the organisation that also produced DALL-E.

There are established real-world applications for AI across industries already, from banking to supply chain, where it can drastically reduce the tasks that people find monotonous, difficult and boring, while simultaneously improving performance.

Computer vision, for example, could certainly help to speed up production in factories, or process thousands of applications in large organisations. We can also expect to see smarter chatbots that are able to respond to our enquiries more accurately, or AI being used to make sense of the masses of data companies now have, including getting rid of junk data. This will speed up the development of accurate and useful AI models.

The idea that AI will replace humans in the workplace is something to be sceptical of. Sure, it can replace certain tasks but there will always be a need for people to oversee the decisions which are being produced from the machine. Further routes of automation seem more realistic as AI begins to become more prevalent in the workplace. This will be highly beneficial to workers as it will cut out monotonous tasks and free up time to focus on more strategic, innovative ones.

The current world of work is shifting and becoming more data orientated. Before we know it, most jobs will require people to have data skills to some degree, or at least be data literate. The availability of low/no code platforms will make it easier for people to get closer to understanding AI even if they don't have coding experience. As people gain more exposure to it, we should see them feeling more confident developing their own insights and potentially statistical models. You don't necessarily need a background in data or maths to get started – you just need to have a play with it and develop an understanding of what's possible and where your limits might be.

Looking ahead, I would expect school children to be taught data skills using the growing number of data science and AI applications available to them, in the same way they're taught literacy and numeracy.

Super computers

One of the reasons why AI has taken off so much in recent years is because of the advances in quantum computing. According to **Moore's Law**, the power of central processing units (CPU) doubles every year – so imagine what we could achieve in 10, 20 or even 50 years if progress continues at that pace. Currently, what holds the training of AI models back are limits in computational power but those barriers are breaking down all the time.

Leaving aside the headline-grabbing stories relating to AI, we can expect to see organisations in more sectors using AI a lot more – for example, in healthcare where tight regulations mean there's still a high degree of trepidation but where the benefits are already clear. Clearly, AI needs to be well-governed, with decisions overseen by humans, so patients' lives aren't put at risk and their outcomes improve.

Data for good

We're already laying the foundations for how ethical AI could support medical research. **SAS is working with Amsterdam UMC** to use AI – including predictive analytics and computer vision – to identify cancer patients who are candidates for lifesaving surgery. The technology allows researchers to collect vast amounts of data, including biomarkers, DNA and genomic data, on patients and make sense of it quickly. As a result, Dr. Geert Kazemier, Professor of Surgery and Director of Surgical Oncology, is confident that 'AI will help us save lives'.

SAS, in partnership with Microsoft, is also supporting vital research into pre-eclampsia – a condition that can cause high blood pressure before and after pregnancy – as part of a pilot at the Institute for Discovery at University College Dublin. The goal of the project is to develop a tool, called AI_PREMie, capable of stratifying risk in patients – so clinicians **can make a more informed decision on diagnosis** and deliver personalised treatments.

As we see more **'data for good'** examples like these, it's likely that public perceptions about AI will change. Over time, it'll become part of the fabric of our lives, as people feel more comfortable using it and consenting for their data to be used – and good governance is key to winning their trust.

WHAT DOES REGULATION LOOK LIKE?

With new regulations to govern AI on the horizon, Kalliopi Spyridaki, Chief Privacy Strategist at SAS, looks at what it means for organisations – and the public.

How exactly AI should be governed has been debated for some time but we're now finally moving towards formal regulation and guidance at a national and supranational level.

The EU has set out plans for **'the first ever legal framework on AI'** – and it is expected to be enshrined in law by the end of 2024, coming into effect a year to 18 months after that.

It has also laid out plans for the **AI Liability Act**, which is designed, among other things, to reduce the burden of proof for victims 'in relation to damages caused by AI systems'. As with the GDPR, I think there will be a campaign to raise public awareness of the protections if these proposals are agreed. This should build trust in the entire ecosystem because consumers know there's enforcement and compensation available in the event of a breach.

In the summer of 2022, the UK government published a policy paper, detailing its vision for a **'pro-innovation approach to regulating AI'**. It's still very early on in the legislative process and it's not yet clear how heavily regulated it will be nor whether, post-Brexit, the UK will see a competitive advantage in taking a different stance to the EU. The US has also produced a blueprint for an **AI Bill of Rights** that includes 'protection from algorithmic discrimination'.

Elsewhere, in Japan, Canada, Australia, New Zealand, Singapore and the Middle East, governments are developing their own frameworks. Some, like New Zealand, are likely to follow the EU's lead, while Singapore has already been active in **providing guidance** and engaging with industry.

Supranationally, organisations like the Council of Europe and OECD (The Organization for Economic Cooperation and Development) are making recommendations around ethical AI and while these won't be legally binding, they will probably influence the laws made in member states.

AI legislation is arguably the most significant piece of data legislation to come after the GDPR and it's easy to draw parallels between the two.

However, the laws around privacy were well established – the GDPR bolstered them and provided greater legal certainty. AI, on the other hand, is still new and complex. Even among EU member states, there may be differences in how the new rules are governed and enforced. Some could set up dedicated supervisory bodies, while others might expand the remit of their data privacy authorities.

Companies will soon have no choice but to abide by the laws in the jurisdictions where they operate. While some of the EU legislation is still being negotiated, there's enough detail for them to start planning on the basis of what's already been set out. Investment is needed to ensure compliance – and this could mean regulators elsewhere align their rules with those of the EU to encourage firms to do business there.

That said, responsibility for ethical AI doesn't lie with policy-makers alone – self-regulation is just as important.

Reggie Townsend, director of the SAS Data Ethics Practice (DEP), believes that anyone who develops and uses AI must prioritise the ethical and social implications in everything they do. This requires everyone to have a greater understanding of how AI works.

He said: "One of my ambitions is to try to make AI less complicated, to demystify it. While the experts know a lot about it, we need to educate the general public. Beyond senior managers, it's important that people understand AI at a level that's appropriate to them because we're all impacted by it."

Bigger companies are already doing a lot around AI ethics – in many cases going beyond what the legislation may require. Microsoft has developed the **Responsible AI by Design** framework, while IBM announced in 2020 that it would withdraw **facial recognition technology** amid concerns about bias. Compliance with any new legislation is non-negotiable but there's a clear business incentive to behave responsibly too.

What SAS is doing

Being at the forefront of AI development, data professionals can also play a leading role in informing policies and developing best practice.

To that end, we created the SAS Data Ethics Practice (DEP), a cross-functional team led by Reggie Townsend. This team leads a globally coordinated effort to ensure that both SAS employees and customers use data-driven systems, including AI, to promote human wellbeing, agency and equity – with the right controls and human oversight to govern the technology.

The team advises and coordinates with SAS Research and Development and other groups within the company to ensure that the key principles of responsible innovation – human-centricity, inclusivity, transparency, accountability, robustness, and privacy and security – are met.

This commitment extends to our customers too. Whether they operate in the public, private or government sector, we want to work with organisations that share our vision for responsible innovation.

Our data ethics practice AI risk management methodology is inspired and informed by a number of responsible AI frameworks, such as **FATE principles and framework** (fairness, accountability, transparency and ethics). This aims to prevent biases, promote responsibility and ensure that models can be explained and understood.

It's important to acknowledge that no single organisation has all the answers when it comes to ethical AI. So we're committed to learning from our partners, customers, academics and the wider industry to build our knowledge and promote best practice, and continually reviewing our projects, products, development methods and contracts.

Click here for more on ethical AI with SAS.

CONCLUSION

Many organisations today find themselves at a crossroads. AI can transform their operations and performance, yet while plenty of uncertainty remains, it could even prompt fears of a dystopian future as portrayed in The Minority Report.

From advancing medical science and tackling climate change, to driving economic growth and creating jobs, there is almost no limit to the benefits AI can bring, as long as it's underpinned by a strong ethical framework. New regulations introduced, in Europe and beyond, may help organisations become more confident in adopting the technology.

We've seen in this report that there are plenty of uncontroversial ways AI can be used to improve business processes but this is only the tip of the iceberg. Responsible experimentation, on a small-scale and within a controlled environment, will drive further innovations over the coming years, without exposing people to undue risk.

All of us have a part to play in ensuring that AI is responsible, ethical and transparent beyond what legislation may require. If needed and when technically feasible, developers should be able to explain how the data is used and how decisions are arrived at to both technical and non-technical stakeholders, and organisations must be able to correctly communicate decision-making to consumers.

No matter how independently AI might appear to act, humans remain firmly in the driving seat. Of course, humans come with their own biases – but high-quality data, transparency and good governance will help to ensure the potential for harm is monitored and reduced.

Choosing a vendor with a human-centric and ethics-led approach to AI enables organisations to navigate the fast-changing landscape, so they can innovate quickly without unwittingly putting people at a disadvantage and losing their trust.

Click **here** to find out more about AI and why it matters