Shaping society 5.0 Building trust in AI as a force for good

AI as a force for good:
realizing the benefits through
openness and transparency

Foreword

**By Harold Pradal,
Chief Commercial Officer, BSI**

1. Signal search analysis 2022-2023.
2. Society 5.0: The Fundamental Concept Of A Human-Centered Society, Open Business Council, Aug 2023
3. BSI Trust in AI Poll 2023

2023 will be viewed as the point that Artificial Intelligence (AI) tipped into the mainstream, with a 286% rise in media coverage of the topic[1]. And whilst headlines were grabbed by ChatGPT, the real AI story is much, much deeper.

This transformational technology is accelerating progress – and has the potential to go further as a force for good and move us towards Society 5.0, a 'human-centered society that balances economic and technological advancement to solve society's problems'[2]. Importantly, it also raises questions around how we build trust in AI and what guardrails are needed to ensure AI shapes our future in a positive way.

In this collection we go behind the headlines to explore the real-world impact of AI through the eyes of BSI experts, drawing on the views of 10,000 people in nine countries. For anyone in doubt, AI is here and it's here to stay – 38% of people use AI in their jobs daily, rising to 70% in China and 64% in India. By 2030, 62% expect their industry will use AI[3].

At BSI we are committed to shaping the impact of technology and innovation for the benefit of individuals, organizations and society. AI sits at the heart of this because it has the potential to be a powerful partner, changing lives and accelerating progress towards a better future and a sustainable world.

We commissioned these essays to turn the spotlight on this generational opportunity – recognizing that the better we understand it, the better we can harness its power. Whether it's creating new workplace opportunities, improving patient outcomes, tackling modern slavery or building a safe global food system, AI has a pivotal role to play.

We examine the importance of embedding digital trust in AI, the critical role for collaboration – between nations, policymakers, organizations and individuals – to unlock AI's true potential, and the fast-evolving regulation designed to ensure consistency and certainty.

With AI crossing over from small, contained environments into mainstream technology at work and at home, this offers a transformational opportunity to unlock a multitude of benefits – provided trust and confidence are present too.

AI is just getting started. At BSI we are excited to partner with our clients as we embark on this journey. We are delighted to present these essays to explore the enormous potential AI offers to shape Society 5.0 and deliver a sustainable future powered by innovation.

AI as a force for good: realizing the benefits through openness and transparency

As a purpose-driven organization, BSI believes AI can be a force for good, changing lives, making a positive impact on society, and accelerating progress towards a sustainable world. In this essay, Craig Civil, Director of Data Science and AI, looks at the steps we can take to help ensure AI can positively shape our future and bring benefits across society.

By Craig Civil

Director of Data Science and AI, BSI

One of the biggest conversations of our time is around AI and how we can use it to change lives for the better, shape the way we work, boost efficiency, and accelerate innovation. From image recognition to linguistic understanding, AI covers a broad spectrum of technologies, which are opening up a vast scope of possibilities for humanity.

In BSI's Trust in AI Poll[1], when asked how they would like to see AI shaping our future by 2050, people around the world prioritized a range of positive impacts, from reducing social inequality (23%) to making it easier for doctors to diagnose medical conditions (28%) to improving education (17%). The world has, understandably, immense hope for how AI can be a force for good.

Nevertheless, the complexity of AI algorithms may leave organizations and individuals feeling overwhelmed by AI's capabilities or concerned about whether to trust AI. AI is not magic, neither need it be mysterious – in fact, it offers the opportunity to drive progress across society in a myriad of positive ways, provided it is effectively managed and well-governed.

Underpinning all AI technology is a plethora of skilled people who have coded it, loaded it with data for training, developed and then tested it with users so that it can serve a credible and useful purpose. The onus to ensure AI is a useful tool that has a positive impact is therefore on both those who are developing the technology and those who are using it.

- As AI technology develops at an ever-increasing pace, having clarity on what AI is (and is not) and how it can be used for good is vital

- Used well, AI has the potential not just to improve society for the many, but to benefit those who are often overlooked and most vulnerable

- Key to using AI for good is ensuring organizations are engaged with and have trust in AI tools by challenging bad data inputs and balancing out implicit biases as well as by sharing openly how they are using both the data and AI tools

28%

of people wanted to see AI make it easier for doctors to diagnose medical conditions

Good data is crucial

Using good quality data, AI can open up vast possibilities. Take the changing climate and protecting the planet as an example, which 29% of people cited as a key area they would like to see AI impacting by 2050. The climate can be considered as one big data system, providing huge amounts of data points every day. AI can gather and analyze this data in a useful and rapid way, identifying, prioritizing, and tracking not only the changes we need to make to mitigate the effects of climate change but also the means to measure and predict the success of an organization's sustainability initiatives.

A good example is The UN Environment Program's World Environment Situation Room[2], launched in 2022 to curate, aggregate and visualize the best available earth observation and sensor data. It is able to perform near real-time analysis and make future predictions based on factors such as CO2 atmospheric concentration, changes in glacier mass and sea level rise.

In healthcare, too, AI is primed to shape our future in a positive way, for example, to assess scans and identify cell anomalies, enabling the highly skilled medic to focus on scans where the AI has identified a potential anomaly. These can then be applied to assessing what the anomaly is and how best to treat it. As a specific example, Google Health, DeepMind, the NHS, Northwestern University and Imperial College, London[3] have already partnered to create an AI model able to spot breast cancer on X-Ray images.

The Covid-19 vaccine is another recent example of where AI was used for good to rapidly synthesize information to model the vaccine and develop new drug interventions. AI algorithms and robotic automation allowed Moderna to move from producing around 30 mRNAs (a molecule fundamental to the development of the vaccine) each month, to around 1000 mRNAs[4].

What unites all these examples is the large volume of high-quality data that is used to train the AI model, working in harmony with skilled human technicians. The old adage 'garbage in, garbage out' holds true, particularly in terms of the data used to create AI models.

29%

of people prioritized AI to help to reduce our impact on the environment and protect the planet by 2050

Transparency can help build greater trust

The magnitude of ways AI can shape our future means we are seeing some degree of hesitation of the unknown. When we asked in our poll about future AI applications, most globally agreed that trust is needed if AI is used. For example, 74% said they need to have trust for AI use in medical devices or treatment and 71% for AI use relating to financial transactions. That is not surprising, given there is the personal information and high, even life-or-death stakes involved, but it emphasizes that to realize the full benefits of AI, transparency and greater communication about its uses is paramount.

Rather than organizations feeling they have to start from scratch to assess the risks of AI tools and mitigate them, guardrails already exist. What is needed – for the public as well as for organizations – is a greater understanding of these checks and balances, and recognition that human involvement will always be needed if we are to make the best use of this technology.

Fear of the unknown could prevent people from adopting AI tools. Greater knowledge of the guidance around it has the potential to free people to make not just good but great use of this technology in every area of life and society.

74%

of people globally said they needed some degree of trust for AI use in medical devices or treatment and 71% said the same for AI relating to financial transactions

Applying standards that move at the pace of change

So, what do we mean by guardrails? Firstly, given AI's global nature, agreed standards and principles of best practice that can evolve alongside the technology and its applications could pave the way to ensure, for example, that data is not misused, and that the inputs being applied to AI tools are fair and equitable. These can include the EU AI Act[5], existing standards that can help manage risk (such as Information technology – Artificial intelligence – Guidance on risk management (ISO/IEC 23984) and the forthcoming AI management standard (ISO/IEC 42001).

There are also guardrails organizations can put in place. Take our own work at BSI as an example. We have a set of Enterprise-wide AI principles that we use to form the backbone of any AI system or model we build in the Innovation team. We adhere to principles that treat everyone fairly and mitigate against the risk of data bias. The ethical use of AI is a key principle that drives our innovation.

On a societal level, having processes in place to identify biases in data that could ultimately lead to a flawed model will be critical to avoiding unintended consequences and ensuring AI is a force for good. A case in point is an AI-led recruitment tool reportedly developed by Amazon[6], that was then scrapped after it appeared that it had trained itself to value male applicants higher than female ones. Having people examine models for the flaws that may emerge is key.

Likewise, the context of our use case requires transparency. At BSI we understand what the code is doing, what model is being created and how it's applying different weightings to the different attributes of the data that is being processed. This means we can provide our human ability to understand and interpret the outputs.

Guardrails already exist. What is needed is a greater understanding of these checks and balances

Craig Civil

Protecting the most vulnerable

It is this human support (sometimes referred to as 'human-in-the-loop') that is a key guardrail. In healthcare, for example, it's striking that in our poll, 57% of people said they supported AI tools being used to treat them – as long as the tools were overseen by a qualified person.

If AI is only as good as the data it is given, people and organizations have an opportunity – and ultimately a responsibility – to ensure the right data is being used. This is where guardrails can help, especially if we want to be sure that vulnerable groups of people are protected from exploitation in the use of AI.

As an example, if organizations use AI to filter job applications for popular roles, it is critical that the process has safeguards and is transparent to its participants. Technology might be able to do the job, but human beings training the AI to do the job, bring with them implicit (and sometimes explicit) biases that can feed through the process. That is, unless the code is rigorously reviewed. A risk-based approach would be aware of the possible flaws in an AI system, for example with facial recognition software, and would then also be able to redress this within the code. AI can be used to correct for institutional biases like this. In fact, with the right data and good governance, AI has the potential to positively transform people's lives.

Going back to Covid for another example, the UK Health Security Agency was able to identify, and advise to shield, an additional 1.7 million people as Clinically Extremely Vulnerable, by feeding technology consultancy BJSS and Oxford University's AI-based risk prediction model QCovid[7] with NHS patient data[8]. While we can hope that this specific situation won't come up again, understanding AI's role in such a life-changing scenario helps bolster confidence in its use, which could be vital for the future.

The opportunity for AI to be a force for good for society is immense. To bring that goodness to fruition requires openness, transparency and trust. As people understand the potential of AI and their power to use it as a tool, embedding guardrails and building greater trust is critical.

57%
of people said they supported AI tools being used to treat them – as long as the tools were overseen by a qualified person

References

1   BSI partnered with Censuswide to survey 10,144 adults across nine markets (Australia, China, France, Germany, India, Japan, Netherlands, UK, and US) between 23rd and 29th August 2023

2   Launch of UN World Environment Situation Room at UNEP@50, UN, March 2022

3   AI breast cancer screening project wins government funding for NHS trial, Imperial College, June 2021

4   Why artificial intelligence is vital in the race to meet the SDGs, WEF, May 2022

5   EU AI Act: first regulation on artificial intelligence, European Parliament, June 2023

6   Amazon scrapped 'sexist AI' tool, BBC, October 2018

7   COVID® risk calculator, QCovid, accessed September 2023

8   Protecting the vulnerable: Data & AI success story, NHS Digital, accessed September 2023

Find out more

BSI's Data Science and AI (DSAI) team is providing new agile ways of innovating with the use of AI. Through collaborative problem-solving with business divisions, clients and tech start-ups, DSAI have identified ways to innovate products and services using data science methods and AI technology. An example is the new algorithmic auditing process, read all about this project here.