Building trust in AI: Ethical framework for a sustainable future

Dr Kevin Macnish

Head of Ethics and Sustainability

Driving *meaningful impact*

Introduction

At the height of the pandemic, pupils were unable to go to school and yet to avoid a breakdown of the education system and the flow of people from school to university, somehow exams needed to continue. The solution the government at the time found was to use teacher assessments as to what pupils would have received had they sat the exam.

Clearly this approach was far from perfect, but the situation was unprecedented. How would the examination boards ensure that favouritism and bias didn't influence the year's A-level and GCSE results? The solution put forward by the Department for Education was an algorithm that would assess each school's past performance to create a predicted average grade, and then correct any significant deviations from that grade.

As soon became apparent, the algorithm used was itself biased, particularly against schools with large class sizes and those which had traditionally underperformed at national exams. This meant that larger schools (typically state schools in cities) were biased against, as were outstanding pupils in traditionally underperforming schools.

It's fair to say that no-one in the system wanted the algorithm to be biased against groups of pupils or against outstanding pupils, and yet it was. How can these both be true? Is it too charitable to assume that the Department for Education didn't want to be biased, or were there rogue data scientists behind the algorithm that secretly wanted to undermine state schools? While we can't know for sure, it was almost certainly neither. It was more likely a case of the Department for Education failing to embed their principles of fairness and non-discrimination into the technology they were building and using. That is, they failed to join up their goals, their values, and their technology. Despite the extreme situation, they should have done better, and paid the price in reputational damage across the press.

That was four years ago. Since then, technology has advanced at pace, and we are now facing tremendous possibilities through developments in artificial intelligence. At the same time, the press, the public and regulators have taken notice of the ethical risks that AI poses. This has led to new legislation such as the EU's AI Act and Californian laws against deep fakes, fines for companies such as Tesla which have over-promised and under-delivered in AI, and reputational harm for organisations which have employed biased AI.

AI therefore introduces both challenges and opportunities that influence how technology is developed, implemented and used. AI ethics includes principles and practices which guide the responsible use of AI. An AI ethics strategy is a means of building those principles and practices into business as usual to ensure that organisational use of AI is responsible.

In this paper, we will explore what an AI ethics strategy is, why every business needs one, and the benefits of having one in place. We'll also consider what should be included in a strategy and how to develop one.

While we all think of ourselves as ethical people, mistakes happen. We may miss key considerations in applying our principles in the workplace as we are pulled in competing directions. Or we may establish principles for ethical behaviour but never flesh out what those principles mean in practice. Creating a robust strategy takes time and expertise in recognising and mitigating the different challenges organisations may face. This is especially true in the case of new technologies, such as AI.

What is *an AI ethics strategy?*

An organisation's AI ethics strategy is a document connecting its vision, values and business plan with its use of technology and surrounding culture. It ensures that the organisation's values are reflected throughout the entirety of the organisation, moving beyond a traditional focus on people's behaviour. The strategy is therefore a clear guide on how ethics will support an organisation's business strategy, and drive its priorities, over the coming years.

For the strategy to be effective, it should be documented in a written format detailing how ethics can assist in achieving business goals. If it isn't written down, it isn't a strategy. Of course, this doesn't mean you can't make changes. In fact, it's easier to adapt and pivot as needed when you have a written document.

Why every organisation needs *an AI ethics strategy*

Every organisation wants to avoid fines for negligence and non-compliance, still more the actual harming of an employee or member of the public. By contrast, the outcomes organisations do aim for are those which make them successful and profitable. As we have seen, though, AI introduces new risks which can threaten that success if not taken into consideration. A well-developed AI ethics strategy will mitigate risks and unwanted outcomes while at the same time accelerate an organisation in the direction it wants to go.

Some of the positive outcomes organisations can see through creating an ethics strategy include:

- Defining an ethics strategy means aligning everything with the organisation’s goals and making sure the strategy works towards achieving them. This approach helps the organisation become proactive, rather than reactive.
- The organisation becomes more efficient as AI is adopted in a more considered way that is consistent with the organisation’s vision, values and business plan. This approach maximises productivity and enables collaboration while mitigating risks.
- The organisation increases its competitive advantage when ethical technology, consistent with its values, helps it deliver better service to its clients.

A strategy ties everything together. For example, it ensures that if an organisation values diversity, this is reflected not just in hiring practices but in the nature of technology that is sought and used. An organisation that claims to value diversity and then purchases technology that is not accessible for some employees or inadvertently biased towards certain skin tones could well face legal and reputational risks.

Key *benefits*

Implementing an AI ethics strategy is not just best practice. It is a critical component for sustainable success. AI ethics strategies can lead to better products and services, brand differentiation, accelerated digital innovation and more.

Better Products and Services

One example of an AI ethics approach is responsible AI, which ensures that ethical principles are at the heart of AI development and implementation. Companies adopting this approach can experience significant performance benefits. Vipin Gopal, Chief Data and Analytics Officer at Eli Lilly, explains: "It would be hard to make the argument that a biased and unfair AI algorithm powers better innovation compared with the alternative. Similar observations can be made with other dimensions of responsible AI, such as security and reliability. In short, responsible AI is a key enabler to ensure that AI-related innovation is meaningful and something that positively benefits society at large."

Research by MIT Sloane and BCG has found that companies that prioritise scaling their responsible AI program over scaling their AI capabilities experience nearly 30% fewer AI failures. And the failures they do have tend to reveal themselves sooner and have significantly less impact on the business and the communities it serves.

Brand Differentiation

In today's competitive landscape, brand differentiation isn't just a strategy- it's a necessity that shapes where and why people choose to work. In the UK, 54% of people choose a place to work based on their beliefs and values. This is greater among younger employees, where 87% of Gen Z professionals would be prepared to quit their jobs to work elsewhere if the values of the new company were more closely aligned to their own, and for 55% a pay rise would not be enough to convince them to stay. On the other hand, 43% of Gen Z employees say that seeing their employers fail to take ESG action was affecting their mental wellbeing. In AI, as early as 2019 younger workers were raising concerns about ethics, with 21% of Millennial employers concerned their companies could use AI unethically, compared to 12% of Gen X and only 6% of Baby Boomers, and many companies clearly do not have (or do not communicate) the policies they have, with 89% saying that, except for AI principles, their company does not have or are unsure if they have specific trustworthy and ethical principles governing emerging tech products. This ties in with a report from the IAPP which found that 80% of companies do not have ethical guidance which extends beyond principles.

Since 2021 there has been a 154% increase in entry-level job postings that mention company values. Hiring replacement staff costs, on average, 120% more than retaining existing employees, with new hires taking on average seven months to break even. Investing in ethics, therefore, can lead to 16:1 return on investment in staffing costs.

73% of consumers are willing to share more personal information if brands are transparent about how the data is being used. However, 45% of consumers would never trust a brand again after it displayed unethical behaviour or is involved in a scandal, and 40% say they'd stop buying from that brand altogether. In 2020 alone, annual spending on ethical products in the UK topped £121bn. It is sobering to read that 92% of employees think that their companies should be doing more to reassure their customers about how personal data is used in their AI.

The active decision to be ethical therefore opens up an organisation to better recruitment and retention, better brand loyalty and better stability. All of these contribute to business success.

Greater digital innovation

Greater digital innovation is key to staying competitive, and an ethics strategy serves as a vital framework for aligning AI advancements with organisational values. Ethics strategies help enterprises focus on innovation by allowing internal and external stakeholders to collaborate on future technology and investment decisions. This collaboration ensures alignment with established organisational values.

Business focus on developing high-value strategies for AI innovation includes initiatives such as:

- Implementation of accessible and non-biased AI.
- Development of AI with adequate guardrails to prevent inadvertent harm.
- New efficiencies of business process through sustainable AI copilots thanks to appropriate risk mitigation.
- Guardrails for innovation to ensure long-term sustainability through anticipating public expectations.

Further Benefits

Other benefits to adopting an AI ethics strategy include:

- Managing existing portfolios by empowering leaders to evaluate current dependencies and associated risk assessments in alignment with the organisation's vision and values. This ensures cost optimisation by conducting a cost-benefit analysis and redirecting funds to high-priority projects to maximise impact.
- Empowering digital transformation by helping organisations progress from early stages to ethical maturity on an ethics maturity index, empowering them to become more adaptive to digital transformation.
- Improving risk management through enabling organisations to identify, assess, and mitigate ethics risks that may pose a threat to their operations. Organisations can enhance their security and ensure business continuity in the face of potential disruptions by developing strategies for mitigating ethics risks. By taking a proactive approach to ethics risk management, organisations can protect their critical assets, maintain compliance with internal and industry regulations, and maintain the trust of their customers and stakeholders.

A good counterexample of this last benefit can be found in the case of Sports Illustrated in which AI had been used to generate content in articles, leading to concerns about transparency and ethical journalism. This controversy damaged the publication's credibility as readers and journalists alike criticised the use of AI without full disclosure.

The risk of not having *an AI ethics strategy*

Even if an organisation has good intentions and values, these might not be represented throughout the entire organisation, and especially in its technology. This can lead to unexpected problems and harm like financial loss, damage to reputation, and even legal issues affecting the organisation, the people involved and its stakeholders. This has become especially apparent with AI, which raises ethical risks relating to bias, privacy, security, trust and more.

These risks could mean fines for negligence and non-compliance of rules and regulation for organisations. Stakeholders, including employees, users and the public, could be harmed by use of an organisation's AI or AI-backed services. All these risk eroding trust in the organisation from regulators, employees, the press and the public. Remember those automobile companies from the last decade!

What's included within *an AI ethics strategy?*

When creating an AI ethics strategy, there are four key elements:

- Foundations,
- Audit of Ethics Infrastructure,
- Desired State, and
- Roadmap.

Foundations

These are the core elements that underpin the whole strategy and can be found in strategic documents at the organisational level. These include:

- Vision and values,
- Business goals, and
- Mission statement.

Audit of Ethics Infrastructure

To develop the strategy, an organisation needs to gather data to accurately assess its current ethical standing. This involves gathering proof about its present ethics, which may differ from its desired or perceived state. The goal is to determine if revising its ethical practices can help achieve business objectives.

An audit for an AI ethics strategy should therefore cover the organisational structure, an overview of technology usage and transactions, and services provided both internally and externally. It should also include governance, ownership of vision and values implementation, and the appointment of C-suite level champions.

A realistic assessment of the organisation's technology and culture is essential. This may be achieved through qualitative methods (such as surveys and interviews) and quantitative methods (e.g. technology audits and policy reviews).

The Audit section should highlight challenges, gaps, and blockers that need to be addressed to reach the desired ethical standards and business goals. The following provides some indicative questions that an audit should cover.

Driving *meaningful impact*

Asset Audit

- What AI are used in the organisation? Are these adjusted to be accessible by default?
- Who is the identified owner or responsible person for each AI asset?

Culture and Values Audit

- What are the culture and values on the ground of the organisation?
- What processes are used throughout the organisation, and who is responsible for each process?

Trust and Transparency Audit

- What are the culture and values of the users/customers of the organisation?
- How do users/customers perceive the organisation?

Digital Literacy Audit

- How literate are employees when it comes to the risks of AI?
- Are employees engaging in best practice when it comes to managing AI?

Governance

- Is there a clear, human-centric governance process in place for AI developed or purchased?
- How is AI employed in governance (e.g. PII-checkers, bias checkers)?

The Audit section should assess the above data and identify the root cause of any problems. This analysis will reveal how effective ethics can drive productivity and offer a competitive edge.

Such reviews allow organisations to link organisational goals and ethics provision. While AI ethics audits may not be your area of expertise or interest, they are fundamental to determine whether ethics can be a strategic tool for identifying performance gaps. Start broad and then go deeper with your audits, as you repeat them.

Desired State

Building on the Foundations and the Audit, a North Star provides a realistic vision for the organisation regarding its technology and surrounding culture. Essentially, it is a "to be" statement as to how each of the elements covered in the Audit will reflect the organisation's values and vision in the future.

Roadmap

The roadmap is a document that outlines how to move from the current state (as seen in the Audit) to the desired future state (the Desired State), typically over 1-3 years. The roadmap should be specific, measurable, achievable, realistic and time-bound, linking directly to KPIs and OKRs for those in charge. The roadmap will detail actions or projects, along with ballpark costs and expected return on investment.

How to develop *an AI ethics strategy*

Strategy helps organisations to focus on how they will succeed. An AI ethics strategy guides organisation on how their vision, values and business plan integrate to help it succeed. There are six phases of a strategic planning process.

Discovery and analysis

Leaders must identify and analyse the gaps and inefficiencies in their organisation's processes, ensuring alignment with the overall vision and values. A discovery project is key to this effort, enabling comprehensive stakeholder feedback through employee surveys, focus groups, and stakeholder interviews. This approach helps to uncover both challenges and opportunities. Gathering data from external agencies, customers, and industry best practices can offer valuable, actionable insights.

Stakeholder buy-in

To successfully implement significant changes, securing stakeholder buy-in is essential. The easiest way to achieve this is by involving stakeholders in the planning process and showing them that the project has the support of leadership. Get stakeholders on the same page with a presentation to senior leadership, followed by a question-and-answer session. Bringing in external AI ethics consultants can also add the required expertise and objectivity, depending on a project's scope.

Identifying and involving the correct stakeholders, including IT, Legal and Risk Management, and Procurement teams in the strategy's development is crucial as their insights, dedication, and responsibility are key to the strategy's success. However, effective stakeholder engagement extends beyond senior members of the organisation. By seeking input from across the organisation, you'll foster broader acceptance and support for the strategy.

Ideally, gathering feedback from external stakeholders - clients, customers and suppliers - ensures the organisation's visions and values align with those it serves. Where customer values differ from those of the organisation, this may lead to a re-evaluation of the organisation's values or its vendor relationships, ensuring a more aligned approach.

Driving *meaningful impact*

Assigning roles and responsibilities

Ethics has a reputation for not being able to speak the language of business, typically lacking a dedicated champion in the C-suite. It risks becoming "everyone's responsibility" which means in effect that it is no-one's. It's essential that a senior figure in the C-suite, with strong interpersonal skills and the ability to educate others about ethics and AI and their ability to fulfil organisational potential, be identified as owner.

Create a detailed project charter outlining the project scope, roles and responsibilities in this phase. Use a RACI matrix to define these roles and increase employee accountability.

Implementation

Clearly outline the budget, deliverables, and timelines for change. Given that external factors may sway the timelines, it is crucial to highlight and prioritise long-term and medium-term goals and objectives. This allows the creation of a plan with a strategic roadmap, connecting strategy with outcomes.

Implementation will need to include a training plan and a communications strategy. These elements can be significant undertakings, requiring a training needs assessment and an assessment of best communications practice. However, these can be incorporated into the Audit process and form part of the initial data collection phase of developing the strategy.

Review and documentation

Not all implementation will be successful first time, and long-term plans need to adapt as circumstances change. Therefore, it is essential to regularly track the progress of your ethics change plan. This ongoing tracking will help project and document it for future reference.

Metrics

The best way to secure future budget is to show the organisation has spent wisely and in a way that enables it to achieve its business goals. The simplest way to do this is through using an ethics maturity assessment. This can be tailored to the specific values and vision of the organisation and will provide a clear, objective measure of success and progress within the organisation.

Annually, aligned with your financial year, review your progress, assess improvements and submit your roadmap and budget for the coming year. Some areas that can provide valuable metrics include:

- Projects: Illustrate their return on investment and show how the adoption of ethics has improved the project.
- Employee Results: Illustrate that staff and clients are satisfied with ethics, using internal or external customer satisfaction scores, such as Great Place to Work (GPTW) or the Net Promoter Score (NPS).
- Customer Results: Use an ethics risk assessment heatmap to illustrate the mitigations in place to cover the most likely and most impactful incidents that could occur. Also conduct regular trust and transparency audits with customers/users to take a temperature check.
- Improvement: Demonstrate through the audits and using an ethics maturity model that standards have been maintained or how they have been improved.

The power of *an AI ethics strategy*

Returning to the A-level debacle of 2020, a large reason for this happening was that there was insufficiently joined-up thinking between the values of the Department for Education and the technology they used in the attempt to remove bias from exam grading. The result was an embarrassing episode for the Department, misery for thousands of pupils, and confusion within the higher education sector at a time when clarity was needed most. Had this happened to a company rather than a government department, at the very least they could have expected a significant drop in the value of their shares and at least as much vitriol from the press as was given to the DfE.

The tragedy is that having the tools in place to avoid situations like this happening is not difficult. As we've seen in this paper, a strategy which ties together organisational plans, organisational values and AI can avert this from happening. Unfortunately, too many organisations see ethics as impacting their people and not their technology. If this was a problem in the past (thinking of the car emissions scandals of the 2010s) then it has only become more so with AI. We will soon reach a point where the risks of organisational use of AI will be sufficiently frequent that where these do result in harms they will lead to negligence cases.

In contrast, as we have seen, by joining up plans, values and AI organisations can benefit massively, from better recruitment and retention, greater customer loyalty, better innovation and better products and services. With an effective AI Ethics Strategy, an organisation will have a shared vision, budget and understanding of where it is going and how ethics will help it get there. The strategy will help integrate organisational values with the technology and culture of that organisation and move ethics and compliance beyond a mere box-ticking exercise. While AI ethics strategies may not be an overnight fix, the potential that they offer organisations to double-down on strengths while mitigating risks is significant. We ignore them at our peril!

sopra steria
next

Get in touch to discuss *getting started with your AI Ethics Strategy today*

kevin.macnish@soprasteria.com

Find out more about our
ethics consultancy practice here.

Driving *meaningful impact*