Harnessing AI for a safer world

© This is Engineering

EngineeringX

Royal Academy of Engineering

LR Foundation

Introduction

From design optimisation to improving process efficiency and predicting maintenance needs, Artificial Intelligence (AI) is rapidly becoming a valuable tool in the field of engineering. Its ability to swiftly analyse vast amounts of data and undertake real-time calculations and predictive modelling can provide valuable insights that help improve human decision-making.

Sectors from agriculture to automotive, energy and healthcare are looking at how they can leverage AI for better economic, social and environmental outcomes. Engineers can use AI tools such as pattern recognition, deep learning, fuzzy logic, and neural networks to enhance sustainability and energy efficiency in construction projects and buildings. Deep generative models can produce new designs based on existing data, offering variations that may have different strengths and weaknesses. AI tools can then analyse design data, compare it to established rules and standards, identify potential issues early in the design process and reduce the likelihood of costly errors or design flaws.[1] AI can perform complex simulations and analysis in robotics, automation and design optimisation,[2] as well as for environmental monitoring, pollution control and resource management.

By applying AI to the right problems, it has the potential to further enhance efficiency, precision and safety in engineering design and processes. As we seek to realise these positive impacts of AI in engineering, we must also manage the risks it presents. This spotlight explores the transformative potential of AI development and adoption while offering practical insights for navigating the challenges from ethical considerations to potential job displacement.

©BP

[1] Algorithm learns to correct 3D printing errors for different parts, materials and systems, University of Cambridge, 16 August 2022, at www.cam.ac.uk/research/news/algorithm-learns-to-correct-3d-printing-errors-for-different-parts-materials-and-systems (retrieved 12 May 2025).

[2] How Is AI Used in Mechanical Engineering?, Massobrio, A., Neural Concept, 7 December 2023, www.neuralconcept.com/post/how-is-ai-used-in-mechanical-engineering#:~:text=Mechanical%20Engineering%20Use%20Case%20%232,exchanger%20designs%20with%20different%20topologies (retrieved 12 May 2025).

What are some of the key challenges associated with AI safety?

While the opportunity to increase safety and efficiency with AI tools is tremendous, it is crucial for all stakeholders to consider AI's limitations when designing solutions. Software engineers, who are at the forefront of AI development, must grapple with ethical and security considerations and communicate AI's limitations to stakeholders and clients. In turn, those adopting AI tools must innovate responsibly to deliver meaningful benefits across the population.

This section explores some of the risks that must be managed to safely harness the full potential of AI.

Equity and bias

There are several ways that the adoption of AI may be inequitable, ranging from the displacement of jobs and concentration of wealth to products and services that do not benefit everyone equally. Unequal access to the positive gains that can be created by AI businesses can contribute to global inequality, with some regions lacking widespread access to the digital infrastructure that powers AI, such as data centres and high-performance computers.

There is a risk that the data sets used to train AI models are geographically or socially biased and not adapted to the regions in which they are applied. This could lead to bias or errors in decisions or outputs and will be particularly visible in areas where limited or no data is available, which may further increase the technological divide.[3] There are, however, techniques to understand and reduce these risks. These include algorithmic assessments or the use of synthetic data, which is designed to address gaps in existing data sets, often because of past bias in the way systems operate or due to scarcity of data.

Privacy and security

Adoption of AI systems and services can introduce new vulnerabilities, both in terms of data privacy and from cyberattacks. Principles such as Secure by Design[4] need to be integrated into the development and adoption of AI systems, particularly for our critical infrastructure and public service delivery. Privacy enhancing technologies and secure data environments can offer extra protection for sensitive data such as personal data or intellectual property, enabling use of data while protecting sensitive information.

Sustainability

Growth in AI and the data centres required for data storage is resulting in increasing demands on critical raw materials, fresh water, and energy while contributing to global carbon emissions.[5] While the use of AI has the potential to deliver environmental benefits, sustainability must be at the forefront in the design and use of new AI products and services to avoid society being locked into systems that are unsustainable in the long term.

[3] "A review of the use of AI in the mining industry: Insights and ethical considerations for multi-objective optimization", Corrigan, C. and Ikonnikova, S. A., The Extractive Industries and Society, Elsevier, 14 Mar. 2024, www.sciencedirect.com/science/article/pii/S2214790X24000388 (retrieved 28 January 2026)..

[4] About Secure by Design, Government Security Group, 6 January 2026 at https://www.security.gov.uk/policy-and-guidance/secure-by-design/about/ (retrieved 28 January 2026).

[5] "As Use of A.I. Soars, So Does the Energy and Water It Requires", Derreby, D., Yale Environment 360, 6 February 2024. https://e360.yale.edu/features/artificial-intelligence-climate-energy-emissions (retrieved 12 May 2025).

What key roles do the industry, government and professional engineers play in harnessing AI for a safer world?

To harness the full potential of AI and ensure everyone benefits from the opportunities it brings, it must be developed and deployed in a way that is safe and ethical both now and in the future. This requires collaboration between professional engineers, government and industry. Each stakeholder group has a vital role to play to ensure AI is developed responsibly based on representative data sets, is subjected to active risk management and robust testing and is applied to the right problems with appropriate monitoring and evaluation.

The following sections highlight the specific skills, capabilities and responsibilities required by each group, emphasising the importance of their collective efforts in shaping the responsible use of AI. Together, these stakeholders can pave the way for a future where AI-driven practices are effective, safe and ethically sound.

Professional engineers

AI developers need to practice active risk management, understanding the limitations of the systems they are developing and identify and address biases that may be introduced through the data or algorithms. This requires a deeper understanding of the range of contexts in which the systems they design are being deployed. Engineers must be able to analyse the data sources, select appropriate models, interpret results, and articulate short- and long-term impacts of their AI solutions on society.

Domain expertise should be combined with an engineering mindset of problem-solving and systems thinking. Technical skills in machine learning, natural language processing, data engineering, programming languages (e.g., Python), libraries (e.g., TensorFlow, PyTorch, scikit-learn), big data technologies, and cloud platforms are crucial for developing and using AI tools effectively. Engineers must ensure that AI is a collaborative tool, requiring diverse multidisciplinary teams to bring together both technical and human expertise.

As AI is an emerging discipline, there are lessons to be shared from wider engineering approaches that design and build safety systems into technologies. A professional ethos among engineers working in AI needs to be developed, embedded and upheld. This should include continuous professional development, systems to report errors or bad practice, and forums to learn from past mistakes, accidents and disasters.

Interviews with academics in the fields of AI and computer science highlighted that a prevalent challenge in teaching AI is the tendency of students to prefer hands-on development over methodology. While hands-on learning in engineering is critical, in AI the desire to learn through active engagement and rapid experimentation, often encapsulated in the 'move fast and break things' approach, can prevent students from gaining formal skills and theory. This can result in attempts to build solutions without a solid foundation in the design principles which are key for embedding security, safety, ethics, and inclusion.[6]

With an increasing number of companies rapidly adopting AI technology as part of their growth strategy, engineers must quickly upskill to help implement these strategies.[7] Educators can also help close the skills gap by strengthening their own skills and understanding of AI and incorporating comprehensive AI education into engineering curricula, ensuring that graduates possess the practical and theoretical skills required to navigate the ever-evolving AI landscape. Some of those needed skills for engineers include:[8]

- **Machine learning:** developing systems that learn and improve from data.
- **Neural networks:** creating models that mimic the brain to recognise patterns.
- **Data analytics:** analysing large data sets to make informed decisions.
- **Predictive maintenance:** using AI to predict equipment failures and optimise maintenance.

These skills enable engineers to leverage AI technologies to create safer, more reliable systems and processes, ultimately contributing to a safer world.[9]

Government

Governments must actively acquire AI skills to understand the technology's implications for their citizens. This can be done through the establishment of specialised delivery units and training programmes or drawing on external expertise through Councils, Taskforces, or expert secondments. In many countries there is a strong push to recruit personnel with AI expertise, as exemplified by the US 2023 Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, which encouraged swift hiring of AI professionals and mandated AI training for employees at all levels. The order was also designed to promote the responsible use and adoption of AI.[10] The UK's A blueprint for modern digital government similarly looks to grow AI capability across government. This includes through updated digital and AI programmes for all levels including senior leadership, expanding the cross-government TechTrack apprenticeship programme, and protecting headcount for digital jobs.[11]

Robust assurance is needed so consumers can be confident that AI systems and services are both safe and trusted. Many governments, including those of Japan, Brazil, Canada, and

[6] Summarized thoughts from five interviews conducted in the fall of 2023 with lecturers in Ireland, the UK, Spain and Denmark.

[7] The Future of Jobs Report 2023, World Economic Forum, 20 April 2023, at https://www.weforum.org/publications/the-future-of-jobs-report-2023/ (retrieved 12 May 2025).

[8] The Impact of AI on the Engineering Field, Johns Hopkins University, 14 June 2024, at https://ep.jhu.edu/news/the-impact-of-ai-on-the-engineering-field/ (retrieved 12 May 2025).

[9] Ibid.

[10] Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, Executive Order 14110, 30 October 2023, at https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence (retrieved 12 May 2025). This order was revoked by President Trump in January 2025 on his return as president of the US.

[11] A blueprint for modern digital government, Department for Science, Innovation & Technology, January 2025, at https://www.gov.uk/government/publications/a-blueprint-for-modern-digital-government/a-blueprint-for-modern-digital-government-html (retrieved 12 May 2025).

the UK, are grappling with AI regulation. The three biggest contributors to the development of AI – China, the US and the EU[12] – have each developed a different strategy of governance and regulation. The Chinese approach focuses on innovation, with the goal for China to lead the world in AI by 2030. This has been accompanied by some of the earliest regulation of recommendation algorithms.[13] The EU has established comprehensive AI legislation with the EU AI Act,[14] which places different obligations on the producer and user depending on the level of risk from AI. To respond to the pace of innovation and opportunities for growth, many countries are developing regulatory sandboxes for AI to create space for companies to test AI tools in a controlled environment under the guidance and oversight of regulators.[15]

As new regulation is developed and comes into force, there has been international regulatory collaboration on AI safety. In November 2023, the UK held the world's first AI Safety Summit,[16] with representatives from 28 countries in attendance. This resulted in the Bletchley Declaration, which represents mutual acknowledgement of the importance of international cooperation on AI safety.[17]

International collaboration has continued through the work of the United Nations Secretary-General's High-level Advisory Body on AI, who published a report in September 2024 setting out proposals for global governance of AI to encourage development and protect human rights, realised through international cooperation.[18] The report also includes recommendations designed to address the global digital and AI divide, such as setting up a global fund and creating a global AI capacity development network.

Alignment on key principles of trustworthy AI is important for international trade and cooperation. Governments play an important role developing international consensus standards for AI to ensure that concepts such as transparency and explainability are understood and integrated into AI systems in a consistent way. Keeping pace with the rapid development of the AI industry creates ongoing challenges for the development and adoption of both standards and regulations. The African Union has developed a Continental Artificial Intelligence Strategy, which calls on member states to adopt a unified approach to ensure consistent and coherent governance, to develop regulatory frameworks that are adaptive to the rapid changes in AI technology, and to collaborate with African nations and other countries to share best practice.[19]

It remains to be seen if governments can effectively enforce regulations and standards

[12] The AI Index 2024 Annual Report Maslej, N., Fattorini, L., Perrault, R., et al, AI Index Steering Committee, Institute for Human-Centered AI, Stanford University, Stanford, CA, April 2024.

[13] An Analysis of China's AI Governance Proposals, Calero, H., Center for Security and Emerging Technology, 12 September 2024, at https://cset.georgetown.edu/article/an-analysis-of-chinas-ai-governance-proposals/ (retrieved 12 May 2025).

[14] EU AI Act: First Regulation on Artificial Intelligence, European Parliament, 8 June 2023, at www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (retrieved 12 May 2025).

[15] Understanding EU's AI Act and its enforcement mechanisms, Barron, W. H., KTS Law. 22 September 2023, at https://ktslaw.com/en/insights/publications/2023/9/understanding%20eus%20ai%20act%20and%20its%20enforcement%20mechanisms (retrieved 12 May 2025).

[16] AI Safety Summit 2023, GOV.UK, 1 November 2023, at https://www.gov.uk/government/topical-events/ai-safety-summit-2023 (retrieved 12 May 2025).

[17] The Bletchley Declaration by Countries Attending the AI Safety Summit, 1-2 November 2023, UK Department for Science, Innovation and Technology, 1 November 2023.

[18] Governing AI for Humanity, United Nations Secretary-General's High-level Advisory Body on AI, 2024, at https://www.un.org/en/ai-advisory-body (retrieved 12 May 2025).

[19] Continental Artificial Intelligence Strategy, African Union, 9 August 2024, at https://au.int/en/documents/20240809/continental-artificial-intelligence-strategy (retrieved 12 May 2025).

meant to protect society from potential harms. Repositories, registration requirements and labels are aimed at providing self-enforcing mechanisms. A more direct mechanism is to adopt steep fines. For example, the EU AI Act bans tools deemed unacceptably risky; failure to comply could result in fines of up to 7% of a company's global revenue or up to €40 million.[20] The act requires all EU member states to create regulatory agencies to implement and enforce the regulation, and enforcement responsibilities belong to the state. This means that enforcement capacity will ultimately vary across the EU.

© This is Engineering

Engineering industry

For AI development many countries are relying on industry to self-regulate. In the US, 15 leading companies have made voluntary commitments to drive safe, secure and trustworthy development of AI. The commitments follow three core tenets: ensuring products are safe before introducing them to the public; building systems that put security first and earning the public's trust.[21] However, without clear national or international regulations, the burden of good governance falls squarely on the shoulders of AI companies.

With the increasing adoption of AI, industry will have to establish and embed new practices. In the EU where AI regulation is in place, companies that use AI systems deemed high risk are required to establish, implement and maintain a risk management system.[22] High-risk AI tools are also required to use data sets that meet specific quality control criteria.

Industry will also have to consider how jobs will change with the adoption of AI and the skills required for AI to be deployed successfully. Historical evidence suggests that while some jobs may be replaced, automation tends to result in a growth in overall employment as it increases efficiency, freeing resources to be reinvested and creating further job opportunities in other areas.[23] AI tools can augment teams by matching skill sets and offering on-the-job

[20] EU AI Act: First Regulation on Artificial Intelligence, European Parliament, 8 June 2023, at www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (retrieved 12 May 2025).

[21] US administration advances AI initiatives, receives further voluntary commitment, Industrial Cyber, 29 July 2024 at https://industrialcyber.co/ai/us-administration-advances-ai-initiatives-receives-further-voluntary-commitment/ (retrieved 27 June 2025).

[22] EU AI Act: First Regulation on Artificial Intelligence, European Parliament, 8 June 2023, at www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence (retrieved 12 May 2025).

[23] No, Robots Aren't Destroying Half of All Jobs, Willcocks, L., LSE Online, at www.lse.ac.uk/study-at-lse/online-learning/insights/no-robots-arent-destroying-half-of-all-jobs (retrieved 12 May 2025).

training to bridge skills gaps.[24] Industries that embrace AI technologies to enhance their workforce rather than replace it are likely to achieve better outcomes. This requires clear governance and accountability for decisions, workplace consultation, and upskilling and reskilling programmes.

Governance and accountability can be enhanced through cooperation between industry and academia. By working with universities, industry can help address the AI skills gap, while also creating graduates with the skills needed for future industry jobs. One example of such a partnership is between Google DeepMind and the African Institute for Mathematical Sciences (AIMS). Google DeepMind is one of the leading AI labs working on projects such as AlphaFold, an artificial intelligence technology that can accurately predict 3D models of protein structures and is being used in biology to address protein folding.[25] They have partnered to create an 'AI for Science' master's programme based in South Africa, designed to develop talent from the continent. Students can pursue advanced studies at AIMS South Africa while accessing a full scholarship and networking opportunities with Google DeepMind's researchers and engineers for mentoring and support. The curriculum was co-created between the partners in consultation with four leading local science institutes, the Square Kilometre Array Observatory, the South African Radio Astronomy Observatory, the Centre for Epidemic Response and Innovation, and the South African Centre for Epidemiological Modelling and Analysis.[26]

Conclusion

The rise of AI-based tools offers immense potential for engineers and engineering to enhance the accuracy, safety and efficiency in design, build and operations. However, it is crucial that a safety culture is embedded in the development and responsible adoption of AI solutions to ensure the risks are understood and mitigated.

To navigate this transformative landscape successfully, a multi-faceted approach is necessary. Government regulations will play a pivotal role in ensuring responsible and safe AI development, implementation and use. Equally important is the cultivation of skills, through both academic education and practical industry experience for professional engineers. Industry investment in AI technology with a focus on safety and reliability will help harness AI technology for a safer world. Key stakeholders must foster a diverse and inclusive environment that embraces a wide range of perspectives to better understand the potential impacts of AI systems and services, and to reduce the potential for bias and inequality. Only with responsible AI adoption that is transparent and trustworthy can we fully harness the potential of AI to transform the engineering industry, driving innovation and create a safer, more sustainable future for all.

[24] Artificial Intelligence (AI) Techniques in Civil Engineering, The Constructor, 9 May 2021, at www.theconstructor.org/artificial-intelligence/artificial-intelligence-techniques-civil-engineering/553331/ (retrieved 12 May 2025).
[25] AlphaFold: Using AI for scientific discovery, Google DeepMind, n.d., at https://deepmind.com/technologies/alphafold/ (retrieved 28 May 2024).
[26] Education Programs, Google DeepMind, 2023, at https://deepmind.google/about/education/ (retrieved 28 May 2023).