RAISE

Responsible GenerativeAI for SMEs in UK and Africa

RAISE Guidelines

31/05/2025

TRILATERAL RESEARCH
Ethical AI

RAi UK

University of Nottingham
UK | CHINA | MALAYSIA

UK Research and Innovation

Contents

Definitions/Glossary

AI – Artificial Intelligence

SEO – Search Engine Optimisation

SMEs – Small and Medium Enterprises

LLM – Large Language Model

1 Why do we need another set of guidelines?

This document aims to provide a simple and usable way for SMEs to navigate the complex world of generative AI and develop sensitivity for possible issues and challenges they need to deal with. It is based on the insight that opportunities and risk arising from generative AI need to be understood in the context of the AI ecosystem within which the SME operates. This means that a successful deployment of generative AI not only requires an understanding of technical risks, but also of the social, economic, legal, and cultural environment.

Based on 12 case studies and a number of workshops with SMEs in the UK and in Africa, we have identified a number of requirements that usable guidelines should fulfil:

- Provide practical guidance
  - Context specific
  - Applicable
  - interactive
- Offer an assessment tool including
  - Risk ranking
  - To-do list
- Provide specific recommendations

This document aims to move in the direction of these requirements and provide SMEs targeted insights and support. This is required for a number of reasons:

1. the lack of control by SMEs over the generative AI tools used, including lack of access to data collection and validation at scale.

2. the developing nature of AI ethics guidance and the lack of its specificity and/or tailoring to the SME research and software development context.
3. the extra effort and time required by SMEs to ensure responsible AI practices.
4. the impact of local culture, norms, and practices on AI use
5. future challenges related to regulatory changes.
6. the pace of developments in generative AI

We provide some conceptual clarification in the following section before we come to the heart of the guidelines, the risk identification strategy in the next section.

1.1 Why worry about generative AI and why now?

In the last decade, generative Artificial Intelligence has kick-started a new wave of AI integration and adoption. These range from systems for text generation, translation, and coding (such as OpenAI's ChatGPT based on Generative Pre-trained Transformer models, Google's Gemini), for image generation (e.g NVIDIA's StyleGAN (Generative Adversarial Network) and Google's DeepDream) and video generation (e.g OpenAI's Sora). The power of these systems has led to a flurry of activities to develop and support them, and they offer significant commercial opportunities in the form of new business models and exploitation. Generative AI systems have the potential to contribute to human flourishing by enhancing creativity and innovation, efficiency, and productivity, facilitating economic growth and sustainability, providing inclusive solutions, and ensuring social responsiveness. At the same time, they raise numerous concerns, including their impact on the educational system, accuracy, intellectual property (e.g., copyright) infringement, data governance, liability, and transparency.

Small to Medium Enterprises (SMEs), including technology start-ups, are driving the development of Generative Artificial Intelligence products and services, whilst some non-technology SMEs are enthusiastically adopting these tools. However, SMEs do not typically have the in-house resources or specialists for responsible AI deployment, particularly in fast-moving areas with uncertain regulation and guidance. Further, SMEs are embedded in commercial ecosystems and therefore can be at the mercy of larger providers. In the case of generative AI, SMEs will often be deploying tools on other parties' terms and conditions.

Beyond the described challenges experienced within the UK generative AI context, SMEs in Africa operate within diverse languages, cultures, and traditions. How generative AI systems align with or are sensitive to socio-cultural needs, contexts and expectations is critically important to the global discourse on ethical AI.

1.2 What are recent developments calling for additional guidance?

'Generative' means able to create. Generative AIs, are software tools that are able to create new content and media, including text, images, video and audio.

It can be thought of as having four core use cases:[1]

- Creation. A user will typically give the AI tool a 'prompt' such as 'create me a job advert for an apprentice chef, to work in a newly opening restaurant' and the tool will use this to generate such an advert.

[1] https://www.mckinsey.com/industries/real-estate/our-insights/generative-ai-can-change-real-estate-but-the-industry-must-change-to-reap-the-benefits

- Concision. A user provides a prompt that links to a text or includes a text, and the Gen AI provides a summary, whether for analysis or as a basis for external-facing text.
- Co-piloting. A system interacts with the user or with the general public in order to guide activities, such as research or customer service routing.
- Coding. The AI provides computer code based on natural language prompts.

The systems can carry out these tasks because they have been trained on a huge number of documents or images and are able to predict what the expected content, format, style, language and tone should be for their outputs. They work best when the desired output is well represented in the data used to train the model.

Some of the big changes with generative AI, when compared to previous AI and data science approaches are:

- **Accessibility** – It is comparatively easy to get started with using generative AI, at least in some manner. Some of the bigger-value uses require specialist expertise, but some tools are simply available to the public via a website or app, or may become integrated with other software on phones, or through other software.
- **Ability to work with unstructured data** – formerly, AI tools needed well collected, carefully processed and formatted data (think of data in a spreadsheet). Generative AI can work with 'unstructured' data such as existing documents designed to be read by people.

- **Ability to ask questions and create content using natural human language** in manner that can be close to having a conversation with a person.
- Potentially allow for **automation of higher level technical and creative fields** (such as creating computer code, business writing or digital images).

1.3 Who are the guidelines for?

We have put together this guidance for small to medium enterprises (SMEs). In the UK this is defined as any organisation with fewer than 250 staff, an annual turnover under € 50m, and a balance sheet total under € 43m[2].It could include everything from a small law firm to a restaurant, to a boutique design firm.

The SME we have in mind:

- Has heard about the potential of generative AI tools but is not an expert.
- Is interested in exploring generative AI tools or adopting them for business use,
- Wants to make sure that they use generative AI in a safe and responsible way that will be best for them, their reputation, their customers and partners, and for wider society.

1.4 Who else might benefit from them?

Many not-for-profit organisations and charities face some of the same concerns as small business (trying to be effective and do the right thing,

[2] https://www.gov.uk/government/publications/beis-small-and-medium-enterprises-sme-action-plan-2022-to-2025/beis-small-and-medium-enterprises-smes-action-plan-2022-to-2025-accessible-webpage#definition-of-an-sme

but with limited resources) and as such may also find this guidance useful.

Some technology start-ups are experts on generative AI. They’re actively developing new generative AI tools and services based around them. These companies might be new and have a small number of employees, but they know generative AI inside out. These companies are not the primary audience for this guidance, but we think they may get something from it too, especially if they have an interest in responsible and ethical use, or their customers include other SMEs.

Many larger organisations are currently trying to understand how to use generative AI and may be looking to do so in a responsible and ethical way. In contrast to SMEs, they are more likely to have existing data protection or compliance teams and dedicated legal and technical expertise. However, generative AI is a new area, and even large organisations may benefit from this guidance.

AI ethics and AI governance are professions that are starting to emerge. We hope this guidance would also be useful for these professionals where they work with or support SMEs.

1.5 What are specific issues for SMEs?

We’re still finding out more about the specific issues for SMEs in adopting generative AI, but we can build on what is known about the technology, the general issues facing SMEs and from other areas of technology adoption and responsible technology use, such as data protection for example. SMEs may find:

- **Lack of specialists and expertise** – a large company might have an R&D department, dedicated data protection lawyers and an AI governance team. An SME is unlikely to have such roles, meaning that AI adoption by SMEs is likely to be driven by generalists, with lots of other things to think about.

- **Lack of resources or tight margins** – SMEs often need to make every penny count.
- Technology, data processing or AI may not be central to their business model. AI might therefore be a **very new area**.
- **Unpredictability** of how an emerging technology will affect each market sector. Analysts make over-arching predictions that generative AI will have profound effects “on the market”. A challenge for SMEs is to adopt and respond to the technology in ways that properly embrace the opportunities that the technology creates, with taking on undue risk.
- **Consumer trust and reputation**: consumer distrust remains a significant challenge due to lack of mechanisms to gain consumer trust generative AI products and services. Maintaining customer trust and safeguarding their reputation are critical for the success of any SME, and the irresponsible use of generative AI can pose a significant threat to both.
- Lack of generative AI standards and frameworks which can be easily translated into standard operating procedures (SOPs) for day to day business operations for SMEs.

On the other hand, SMEs can be great innovators with the drive to try out new ways of doing things. This is an area where we really want to check our assumptions.

1.6 SMEs in the UK and Africa

Small and medium enterprises (SMEs) in both the United Kingdom and Africa are increasingly embracing artificial intelligence (AI) technologies, including generative AI, to drive innovation, enhance productivity, and gain a competitive edge in their respective markets.

However, as these SMEs integrate generative AI into their operations, they face unique challenges and considerations that necessitate clear guidelines for responsible use.

A major reason SMEs in both regions require guidelines for responsible AI use is to navigate the ethical implications of generative AI. Generative AI has the capability to produce highly realistic content, such as images, text, and music, autonomously. While this technology presents numerous opportunities for SMEs, it also raises concerns about ethical dilemmas, including potential misuse, manipulation, and infringement of intellectual property rights. Clear guidelines can help SMEs manage the ethical implications of using generative AI and provide them with practical ways to ensure transparency, accountability, and fairness in their AI applications.

Also, guidelines for responsible AI use are essential for SMEs in the UK and Africa to address regulatory requirements and compliance challenges. Both regions have distinct legal frameworks and regulations governing the use of AI technologies, such as data protection laws and intellectual property rights regulations. Many African countries are in the process of developing or updating their regulatory frameworks for AI technologies, including data protection laws, intellectual property rights regulations, and guidelines for AI ethics and governance. SMEs must navigate these complex regulatory landscapes to avoid legal pitfalls and ensure their AI deployments are compliant with local and international standards. By providing SMEs with clear guidance, these can help mitigate legal risks and foster a culture of responsible AI adoption.

Furthermore, guidelines for responsible AI use can empower SMEs to build trust with customers, partners, and stakeholders. By demonstrating a commitment to ethical AI practices SMEs can enhance their reputation and credibility in the marketplace. This trust not only strengthens relationships with existing stakeholders but also opens doors to new business opportunities and partnerships, driving long-term success and sustainability for SMEs in both the UK and Africa.

Therefore, by providing clear guidance on ethical considerations, regulatory compliance, and building trust, these guidelines can empower SMEs to harness the transformative potential of generative AI while mitigating risks and contributing to the responsible development and deployment of AI technologies in their respective regions.

1.7 Opportunities

This document focuses on the risks that can arise from generative AI. However, it is clear that generative AI technologies also offer significant opportunities. There is a potential to enhance creativity and innovation through automated idea generation, prototyping, efficient data analysis for market trends, and creative content generation. Responsible integration of generative AI systems can also significantly increase efficiency and productivity. This can be through optimised resource allocation (e.g inventory management), improved data analysis and automated repetitive tasks. Another opportunity for SMEs is that it can help to personalise customer experiences. By understanding individual customer behaviour through analysis of vast amounts of data, SMEs can tailor their services to meet specific customer needs. Integration of generative AI systems can also play a crucial role in facilitating market expansion for SMEs by providing relevant market research, predictive analytics for target identification, supply chain optimisation and language translation. It can also assist in risk mitigation and management. Generative AI system driven tools can be developed for continuous monitoring of market conditions as well as internal and external factors that may pose risks. Such systems can predict potential challenges, identify vulnerabilities, and support crisis response planning. In essence, the responsible integration of generative AI systems can differentiate SMEs in the market, build trust with customers and stakeholders and thus can help gain competitive advantage.

2 Identifying risks and priorities

Thinking about AI in terms of risks and opportunities helps SMEs put AI into a perspective they are used to and apply risk mitigation. Feedback from SMEs indicated that they were interested in specific and practicable input that allows them to plan their intervention and to draw on existing knowledge and expertise. They were interested in an interactive tool that is easy to use, interactive and context specific. This tool, the RAISE SME tool, is in the process of being developed as an interactive web-based tool that will initially be made available on the RAISE project website.

The following figure 1 gives an overview of the logical flow of such a tool whose content will be described in more detail below.

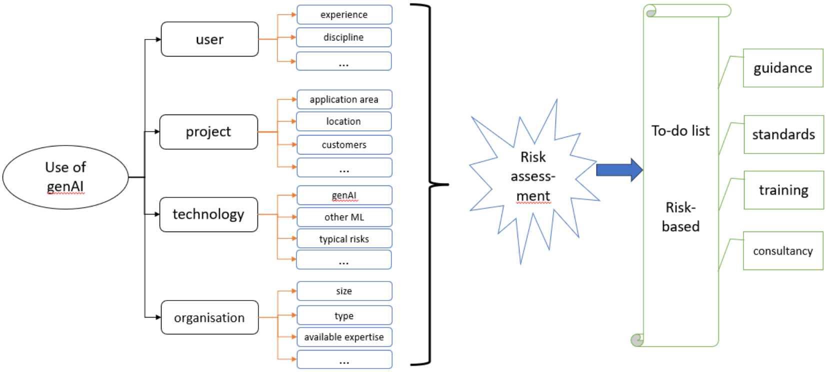


Figure 1: schematic overview of the RAISE SME tool

The figure starts with the SME’s decision to use generative AI. The RAISE tool then collects information about the planned use of generative AI that is relevant to identify pertinent risks. Questions about the user, project, technology to be used, and organisation are needed to identify specific risks. These specific risks and generic risk linked to generative AI are then self-assessed by the user of the tool, taking into consideration their confidence in dealing with them. This will provide the basis for the ranking of the pertinent risks. The ranked risks will be used to generate a to-do list for the system user. The to-do list will link to proven mitigation strategies, thus allowing the user to not only understand the risk but also to find ways of addressing them.

The majority of the remainder of this document provides detail on the content and structure of this RAISE SME tool. It can thus be read as a requirements document and content provision for the tool. The text includes explanations of the tool as well as text to be used in the tool. Text to be used verbatim in the tool is shown in italics.

2.1 Risk assessment and ranking

The first section of the tool asks multiple-choice type questions about how well the SME understands risks, how well they believe to have covered them and how confident they are in their answer. This risk assessment part covers all generic risks as well as the specific risks that are identified on the basis of the screening questions (e.g. based on application area, data types, or technology used).

The principle here is that for each likely relevant risk the SME needs to understand how likely it is to materialise and how significant its impact will be. This traditional risk score is moderated by the mitigation score which is calculated using the SME's understanding of the risk and existing mitigation strategies. For each risk the user will be asked to respond to 4 Lickert scale questions. The first two of these are used to calculate the risk score, the remaining two are used to calculate the mitigation score. The following example uses the generic risk of bias to demonstrate the logic of data collection by the tool.

Please indicate below the degree to which you agree with the statements.

Bias				
a) Bias in our generative AI system would have a significant impact in terms of number of people and / or significance on				
Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
b) It is likely that bias will manifest in our generative AI system				
Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
c) We are aware of and understand in which way bias may be integrated in our system or may arise from our system.				
Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
d) We have appropriate measures in place to understand bias in our system and mitigate its consequences.				
Strongly disagree	*Disagree*	*Neutral*	*Agree*	*Strongly agree*
***Bias** – AI systems replicate patterns in the data they are trained upon. If this data is incomplete in some way, for example, if more is known about some parts of the population than others, then the AI systems can have different results for different groups. AI can also replicate historical patterns of unequal treatment.*				

The overall risk score will be calculated as follows:

Lickert scale answers are translated into scores from Strongly Disagree = 1 to Strongly Agree = 5.

Risk score: Impact x likelihood or (a) x (b)

Mitigation score: awareness x existing mitigation or (c) x (d)

Overall risk factor (RF) = risk score – mitigation score

RF = (a) x (b) – (c) x (d)

Rationale: the first part of the equation is the traditional calculation of risk, namely the probability of the risk materialising multiplied by the severity of the impact (in terms of scope and scale).

The second part of the equation is what we call the mitigation score. This is calculated by your level of understanding of the problem multiplied by the quality of the mitigation measures you already have in place.

This method of calculating the risk will allow the tool to identify and prioritise those risks which have not yet mitigated. Risks can be ordered according to mitigation score.

2.2 Screening questions

The initial task of the RAISE SME tool is to gather detail about the planned use of generative AI which is required to identify specific risks it may arise. This is done through a series of screening questions.

The following questions explore the system you are planning or using and its context to provide input into the evaluation or the risk elicitation in the following steps. This may be important, as context may inform the relevance of risks. For example, data protection questions will typically be more important in a healthcare environment than in a gaming environment.

- *What is the application area or sector that you want to use it in?*
    - *Healthcare*
    - *Transport*
    - *Art*

    - *Journalism*
    - *…..*
- *Geographical location: Where are you based?*
    - *UK*
    - *Africa*
    - *...*
- *What types of data (e.g., customer data, internal documents, publicly available information) will your SME be inputting into the AI system*
    - *Personal data*
    - *Special category data*
    - *Commercially confidential data*
    - *Publicly available data*
- *Who will be your customers?*
    - *Consumers*
    - *Businesses*
    - *Government*
- *What is your funding mechanism*
    - *self-funded, bootstrapped*
    - *grant funded*
    - *venture capital*

2.3 Generic risks and specific risks

The scoring of the risk factors for all relevant risks requires an understanding of which risks need to be covered. This section lists both the generic risks, which are likely to be relevant to all uses of generative AI as well as the specific risks. The RAISE SME tool will cover the generic risks for all users and the specific ones depending on their responses to the screening questions.

2.3.1 Generic risks of generative AI

The use of generative AI leads to certain risks which are widely discussed in the literature and that may arise independent of specific application area.

Risk category	Specific risks
Technical risks	• Transparency • Explainability
Organisational risks	• Organisational Preparedness • Financial Risks • Shadow AI • Employee Satisfaction • Platform Risks
Compliance risks	• IP Infringement • Dual Use • Data Protection
Societal and cultural risks	• Global Labor Conditions • Sustainability • Market Competition • Localisation and Cultural Context

2.3.2 Specific risks of generative AI (application)

Risk category	Specific risks
Healthcare	• Risk of pricing out underserved users/patients. • Potential liability in cases where AI-generated responses lead to misdiagnosis or harm • Inadequate insurance coverage for bodily harm or injury caused by AI decisions

	• Cultural mismatches between AI responses and local health beliefs or practices • Failure to provide adequate emotional or follow-up support based on AI interactions • Weak integration between AI outputs and medical workflows (e.g., doctor follow-ups). • interoperability with existing healthcare systems (e.g., EHRs)
Journalism	• Cultural and Linguistic Inaccuracy (difficulty in pronouncing indigenous words) • Public misunderstanding and acceptance (due to lack of understanding of AI avatars and their limitations) • High subscription costs for voice cloning • Cloning voice quality • Limited options in local avatar appearance and customisations
Art	• Inappropriate or Harmful Content Generation • Copyright and Intellectual Property Violations (The AI may unintentionally replicate or mimic the style of existing artists) • Legal uncertainties regarding which data can be lawfully used for

	training models, especially around scraped or publicly available art. - Risk of app being blacklisted by major application platform providers e.g google playstore
Legal (Lease)	- Risk of sensitive client data being inadvertently used in model training or exposed via outputs - Risk of generating legally incorrect, misleading, or fabricated information that could affect legal decisions. - Difficulty in understanding or tracing how decisions are made by the model, especially in high-stakes legal contexts. - Relying on internal contractual agreements rather regulatory authorities for guidance due to lack of legal AI regulations
Financial investment	- Allowing a generative AI model to provide financial advice may violate financial regulatory requirements - Users may misinterpret the chatbot's educational information as certified financial advice - Obtaining financial investment licenses or certifications for financial AI systems is challenging thereby reducing users trust in the platform

Corporate Intelligence	• Risk of Data Quality Issues from AI Processing of data about corporations • Regulatory compliance risks due to corporations existing under different jurisdiction • AI tools may inherit or amplify biases in training data, especially when extracting or summarising information from unstructured sources used to provide corporate intelligence.
Education and Corporate Training	• Misalignment with National Education Policies: AI-generated content may not always align with national curriculum standards. • Using multiple models without unified benchmarks or adaptation protocols may yield inconsistent quality across subjects or domains. • Reliance on external providers may pose risks related to service continuity, vendor lock-in, or changes in licensing • Ambiguity around update schedules, data handling, and system limitations could lead to trust issues among users (e.g., students, educators, institutions). • formal quality assurance or regulatory oversight risks • Users, especially minors or under-resourced learners, may not fully

	understand what they are consenting to regarding data use and AI engagement • Collecting and processing student learning behaviors or performance data via AI tools without adequate safeguards could lead to exploitation or surveillance concerns. • Efforts to democratise AI may still be limited by digital divides (connectivity, hardware, language) • Models presented as “free” or “affordable” may include embedded costs (ads, data extraction, limited features) that undermine equitable access.
Platform Provider (decentralised agent based)	• Support of peer-to-peer AI agent communication, including subcontracting tasks to other agents could lead to untraceable or opaque decision-making chains, increasing the risk of irresponsible or harmful outputs, especially when generative models are involved. • Risk of Malicious AI Services: since operators can upload any kind of AI container (AIM)—open or closed—without restrictions, there's a risk of malicious or biased AI agents being deployed onto the network with little oversight

	• Risks arising from the absence of user facing policy documents that define acceptable use, liability, or dispute resolution mechanisms • Decentralized governance and jurisdictional complexity risks particularly when agents exchange data or services across borders. • Unclear accountability for harm. who is responsible when an agent behaves in an unintended or harmful way—especially in a permissionless ecosystem • Reliance on external blockchain infrastructure such as Ethereum for token and license management can create scalability issues, gas fee volatility, and external dependencies. • Scalability risks around end user interface, payment for compute, security, uptime and equitable compute distribution. • Lack of content moderation: Being agnostic about the content and models used by node operators increases the likelihood that harmful or unethical AI applications (e.g., biased models, misinformation generators) could proliferate without sufficient oversight. • Social and economic equity risks.

3 Risk descriptions

3.1 Generic Risks

3.1.1 Technical risks

Bias – AI systems replicate patterns in the data they are trained upon. If this data is incomplete in some way, for example, if more is known about some parts of the population than others, then the AI systems can have different results for different groups. AI can also replicate historical patterns of unequal treatment.

Explainability poses a significant risk for SMEs adopting generative AI, as the complex and opaque nature of these models can make it difficult to understand how outputs are generated. This lack of transparency can undermine trust in the technology, hinder regulatory compliance (especially under data protection and AI governance frameworks), and increase exposure to reputational or legal risks if decisions based on AI outputs cannot be justified. For SMEs without in-house AI expertise, this challenge is compounded by limited resources to audit or interpret model behaviour, potentially making explainability a barrier to safe and effective AI adoption

Transparency is a critical risk for SMEs implementing generative AI, as these models often operate as "black boxes," making it difficult to trace how data inputs lead to specific outputs. This opacity can hinder accountability, making it challenging to explain decisions to customers, partners, or regulators—especially in sensitive domains such as finance, healthcare, or recruitment. A lack of transparency can also erode stakeholder trust and limit the SME's ability to detect and correct errors or biases in the system. Without clear insight into how the AI functions, SMEs may struggle to align their use of generative AI with ethical standards and regulatory expectations.

3.1.2 Organisational risks

Technology reliance and deskilling - where people come to depend upon an AI system to do part of their job and become less able to understand when a system has gone wrong or has failed. They may take the results of AI systems as unquestionable true and never think to challenge or check them.

Financial risks are a major concern for SMEs adopting generative AI, as the development, integration, and maintenance of these systems can require substantial investment in infrastructure, expertise, and ongoing support. Unexpected costs may arise from licensing, cloud computing, or the need for specialized personnel to manage and monitor AI tools. Moreover, poor performance, biased outputs, or non-compliance with legal and ethical standards could result in financial penalties, lost business opportunities, or damage to reputation. For SMEs with limited financial buffers, such risks can quickly outweigh the potential benefits of generative AI adoption. **Error based on poor quality training data** – sometimes, there is just not enough training data available on topic or issue to properly train an AI. This particularly the case for rare or unusual events. This means that these areas can be more prone to error or mistakes.

Shadow AI is a term given to the use of AI tools by employees without the awareness of senior management or leadership, often in ways which may violate more formal policies on the use of the AI that a company may have adopted.[3] Examples would include an employee

[3] https://www.forbes.com/sites/delltechnologies/2023/10/31/what-is-shadow-ai-and-what-can-it-do-about-it/

putting customer details into ChatGPT through a web browser or on their phone and asking it to generate an email to them.

Platform risks: Many gen AI tools offer access to the software running on the providers’ servers, typically through an API or a web interface. This makes it easy to start using these tools and limits the start-up cost involved in doing so. Some companies are even building business models around offering an easy to use or specialist interface on top of a generative AI platform. However, this model does leads to risks for business that become reliant on these tools. For example, API access could be cut off, be unreliable, or the price could suddenly increase. The end user is somewhat at the mercy of the provider.

Employee satisfaction: Generative AI has the potential to significantly impact employee satisfaction and feelings of job security, with both positive and negative consequences. While automation driven by generative AI can streamline processes and enhance efficiency, it may also lead to job displacement, skill redundancy, and increased workload or stress for employees. Moreover, the introduction of AI technologies can undermine employees' sense of control or autonomy over their work and raise ethical concerns related to content generation and organisational values. Additionally, the erosion of trust and morale within the workplace may occur if employees perceive management's decision to implement AI as prioritizing efficiency over human well-being.

Damage to brand or reputation: Companies with brands based around authenticity, a personal touch, human contact might experience customer push-back around their use of generative AI which might be seen as conflicting with these values.

3.1.3 Compliance risks

Perceived or actual **IP infringement**, inappropriate mimicry - generative AI are typically trained on very large numbers of images or texts. The legal status of these is not always clear. The outputs of generative AI may potentially infringe on copyright and intellectual properly.

Dual use - where AI developed for civilian applications can also be used for military applications. This may lead to infringement of legislation, notably export control restrictions

Data protection: Users can provide data to generative AI tools and ask the tool to interpret, summarise or analyse this data, or to use this data in the product of some other output. This is quite a powerful tool, but SMEs need to consider if they should be sharing this data with the organisation providing the generative AI (which they might be doing if they upload it to a tool). It might be personal data, or commercially confidential.

3.1.4 Societal and cultural risks

Global labour conditions,[4] **sustainability**.[5] AI systems can be labour and resource intensive. Some have used low-paid labour to train AI models, and the process can use large amounts of energy, water and rare materials.

Errors arising from misuse.[6] - AI can just be used incorrectly, or for different purposes than it was created for. Whilst generative AI seems

[4] https://time.com/6247678/openai-chatgpt-kenya-workers/

[5] https://www.accenture.com/il-en/blogs/consulting/making-generative-ai-green

[6] See, e.g., https://www.theguardian.com/technology/2024/jan/20/dpd-ai-chatbot-swears-calls-itself-useless-and-criticises-firm

to have a very wide range of uses, we are still learning what it is good for, and what it is not.

3.2 Specific Risks

Health

The integration of generative AI into healthcare platforms presents a complex array of risks spanning financial, legal, technical, ethical, and sociocultural dimensions. Financial constraints, particularly in resource limited settings, directly impact the ability to invest in secure infrastructure, regulatory compliance, and skilled personnel, thereby compounding legal and operational vulnerabilities. The absence of clear liability frameworks for AI-generated medical advice further heightens exposure to legal risk, especially in the event of harm. Technical challenges related to data quality, algorithmic bias, and system integration threaten the reliability and safety of AI outputs, while limited digital literacy and language exclusion risk marginalising already underserved populations. Moreover, the lack of emotional intelligence in AI systems raises ethical concerns, particularly in contexts involving mental health or distress. Collectively, these risks underscore the urgent need for robust mitigation strategies, contextual regulation, and inclusive design in the deployment of generative AI in healthcare.

Art

The use of generative AI in the creation of art introduces a range of ethical, legal, and social risks. One of the most pressing concerns is the generation of inappropriate or harmful content, such as sexually explicit images, especially when models are trained on datasets containing nude or sensitive artistic works. Additionally, generative AI raises complex copyright issues, as it can inadvertently replicate protected artistic styles or content without permission, blurring the lines between inspiration and infringement. These challenges are compounded by broader concerns about artistic authenticity and the potential devaluation of human creativity, as well as legal uncertainties surrounding the use of publicly available or scraped data for model training. Platforms that host or promote AI-generated art must also navigate reputational risks stemming from misuse by users and the high resource demands of these technologies. As the boundaries of creativity are redefined by AI, it is crucial to develop responsible frameworks that safeguard artistic integrity while supporting innovation.

Journalism

The use of generative AI for journalism, media production offers notable efficiencies but also presents several risks that warrant careful consideration. Foremost among these is the potential for job displacement, with journalists expressing concern over being replaced by AI-generated content. Cultural and linguistic inaccuracies also pose significant challenges, particularly in African contexts where AI tools often struggle to pronounce indigenous words correctly, leading to diminished authenticity and audience engagement. Furthermore, public misunderstanding of AI-generated avatars perceived as lacking emotional depth and humour has led to scepticism and resistance from viewers. Ethical concerns surrounding transparency in content creation have emerged, necessitating explicit disclosure to maintain trust. Financial constraints further compound these issues, with the high cost of advanced AI features such as voice cloning limiting equitable access and adoption. Collectively, these risks underscore the need for culturally responsive, transparent, and economically accessible AI solutions in the media sector.

Education and Training

Key concerns in the use of generative AI in education and training include the potential for bias and inaccuracy in AI-generated content, especially where such outputs influence learning outcomes or replicate systemic inequalities. The use of third-party models raises further issues around data privacy, model transparency, and infrastructural dependency, particularly when dealing with sensitive or government-issued educational materials. Moreover, the opacity of generative AI systems can challenge explainability, accountability, and user trust, while increasing the risk of security vulnerabilities through API integration. Without robust mitigation measures such as expert oversight, transparent data practices, and security protocols these risks may compromise the intended benefits of generative AI in enhancing educational delivery and equity.

Legal

The integration of generative AI into legal technology solutions especially in drafting legal contracts and leases presents a set of risks that span ethical, technical and operational dimensions. Among the most significant concerns are data privacy breaches, AI hallucinations, and overreliance on or misinterpretation of AI outputs by users. Additional risks include the opacity of AI decision-making, inadequate organisational governance, the selection of inappropriate models, and exposure to harmful or non-compliant content. These issues are compounded by the dynamic nature of both AI technologies and regulatory expectations, which renders static mitigation strategies insufficient over time and necessitates continual reassessment and adaptation.

Financial investment

SMES offering financial investment services can encounter several generative AI related risks. A key consideration is the regulatory risk associated with offering advice through AI systems, which may inadvertently breach compliance requirements if not properly constrained. There is also a significant ethical concern that users could mistake educational content generated by AI for formal financial advice, potentially leading to poor decision-making. Technical and infrastructural challenges, such as integrating AI across multiple customer engagement platforms like WhatsApp, add complexity, particularly when multiple bots operate concurrently. Moreover, issues related to data quality and training bias can undermine the effectiveness and inclusivity of AI responses, while the high cost of deploying proprietary models raises questions about financial sustainability. Concerns around user trust and frustration also emerge, especially when AI is used in emotionally sensitive customer support scenarios. Finally, the use of AI in handling personal and financial data brings data security and privacy considerations to the fore, underscoring the need for robust governance frameworks.

Corporate Intelligence

In the field of corporate intelligence, the automation of tasks like data extraction, translation, and transcription may also lead to the loss of contextual nuances. Without formal governance structures, SMEs risk deploying these tools without adequate oversight, increasing the likelihood of ethical lapses, biased outputs, or regulatory non-compliance especially around the different jurisdiction where corporations exist and where their data resides. Furthermore, access to high-quality models is often limited by cost and technical barriers, exacerbating inequalities in adoption and innovation. These risks highlight the urgent need for responsible AI practices tailored to the constraints and realities of smaller organisations.

Platform Provider (decentralised agent based)

The implementation of generative AI within decentralised platforms presents a couple of risks that span technical, ethical, regulatory, and socio-economic domains. Key concerns include the potential monopolisation of AI capabilities by large technology firms, the infringement of copyright through the use of unlicensed training data, and the inherent difficulty of regulating AI behaviour in a decentralised ecosystem. Furthermore, challenges such as limited technical readiness, data security vulnerabilities, legal uncertainty, and the risk of talent poaching threaten the sustainable growth of such platforms. These risks are compounded by the economic fragility of emerging ecosystems and the potential overcomplexity of coordinating diverse stakeholder interests. Addressing these concerns requires robust mitigation strategies that prioritise openness, responsible governance, infrastructure support, and community-led oversight.

3.2.1 Global

Some of the challenges that SMEs face globally include but are not limited to difficulties of anticipating commercial opportunities accurately; high cost of implementation, lack of availability of data, complex regulatory landscape, limited technical and human infrastructure for implementing interoperable systems, and security risks. Others include difficulties of ensuring data privacy and appropriately determining intellectual property rights. Ensuring that these systems are scalable with the growth of the business is also a critical challenge. SMEs may also often struggle to clearly quantify the return on investment for responsible integration or have employee resistance/lack of understanding of the technology. Finally, there is the challenge of ensuring this integration is environmentally friendly since training generative AI models and inference processing have significant carbon footprints.

3.2.2 UK and Africa

In addition to the above challenges identified as global, there are also challenges SMEs in the UK and Africa can face. These include:

- Market Competition
    - The presence of other enterprises working similar integration initiatives (including larger ones) with more financial resources
- Exploitation of labour
    - The tendency to exploit cheap labour in Africa
- Localisation and Cultural Context
    - Difficulty of localising services considering diverse cultural contexts
- Data Availability and Quality
    - Often lack of data or quality data especially in Africa
- Access to Skilled Talent
    - Limited human infrastructure for responsible integration

Some of the challenges highlighted in this sub section (UK and Africa) showcase a blend of contextual nuances and global relevance.

3.3 What might SMEs do wrong with generative AI?

There are irresponsible practices that some SMEs might want to experiment with. Avoiding these would be a good part of shifting towards the responsible adoption of generative AI.

- **Replacing a human-led process with generative AI output, when there is a legal responsibility to do the process.**
    - There are some processes where a person has a duty to think about something and come to a conclusion (for example a risk assessment, health and safety assessment, or a data protection impact assessment). There’s often a documentation of this process, and businesses might be tempted to use generative AI as a short-cut to create this documentation, but they would not be fulfilling the legal requirement itself.
- **Replacing internal practices or processes when generative AI does not have the real capability to provide that process.**
    - Businesses might be driven to reduce costs, reducing time spent on a task or even cover a lack of specialist expertise, by using generative AI for tasks where it is not competent, perhaps because of misunderstanding the capabilities of the tools.
    - For example, providing feedback to employees, creating contracts, creating internal processes, writing communications with staff etc, or any area where the output might seem plausible, but is not actually appropriate for the company or the context of use.
    - e.g. asking ChatGPT to generate a contract for a particular customer. This will likely produce a plausible looking contract, that might be *good enough* for some circumstances, but will not reflect specialist needs of the parties, or really understand the deal being made.
- **Not checking on content generated by generative AI.**
    - SMEs will be tempted to use Generative AI to produce marketing and promotional content. However,

Generative AI is prone to ‘hallucinations’ - making statements of fact which are not true. This can potentially expose companies to risks around discriminatory speech or images (e.g., racism, sexism, ablism, ageism, homophobia etc), but also to libellous or misleading content. Content generated by generative AI might also be dangerous by giving unsafe advice (e.g., DIY, health, exercise, legal advice, and many other topics).

  - Similarly, Generative AI is increasingly being included in search tools, to provide summaries of information. SMEs may find this a useful way to find useful information, but ‘hallucinations’ in this context can include false or untrue information.

- **Not training staff or setting expectations** about how generative AI tools should be used in the workplace.
  - Some generative AI tools are easily accessible, and employees might start using them on their own. Employers should be clear with staff about what are appropriate or inappropriate uses.
  - On the other hand, because of the way that generative AI is more accessible than earlier AI, employees can become a good source of ideas on how to make use of generative AI in the workplace.
- **Ignoring generative AI**
  - SMEs might simply assume that generative AI is not important to them and that they do not need to think about it.
  - They might suffer a competitive disadvantage in relation to other companies who can make responsible use of the technology.

- It leaves them open to the risk of ‘shadow AI’ - where their employees adopt generative AI tools without permission.

In February 2024 The "Glasgow Willy Wonka Experience," touted as a whimsical journey through a chocolate factory akin to the beloved Willy Wonka tale, ended up as a letdown for attendees. Despite assurances of magic, the reality fell short with a drab venue, cheap props, and an AI-generated script that failed to capture the essence of the story. Misleading advertisements using AI-generated imagery left disappointed children and fuelled internet memes. Priced at £35 per ticket, outraged parents sought refunds and involved the police, leading to the event's premature closure. The disenchanted Oompa-Loompa and the perplexing "Unknown" character only added to the confusion, highlighting the risks of relying solely on AI for event planning and the importance of human oversight and creativity.

4 Mitigation Strategies

4.1

To mitigate the above risks and challenges, it is important for SMEs to use responsible research and innovation (RRI) tools for effective identification of risks and responsible actions. For instance, the AREA (anticipate, reflect, engage and act) framework can help the SME to; anticipate the potential social, economic, and ethical impacts of integrating generative AI systems; reflect on the ethical implications of deploying generative AI systems, engage with relevant stakeholders throughout the AI integration process and act by implementing clear ethical processes, actively monitor the systems’ performance and impact, and act in accordance with relevant legal and regulatory requirements. The following table attempts to provide some practical actions that mitigate the risks identified in the previous sections

Specific Risks	**Mitigation Strategies**
Bias	• have a clear understanding of how the AI model was trained, including the datasets used and the potential biases they might contain. • Depending on the type of data (e.g., customer data, internal documents, publicly available information) that your SME will be inputting into the AI system identify any underrepresented data categories or potential biases • Consider setting up a bias Audit Unit (https://www.holisticai.com/blog/unconscious-ai-bias-audit-your-company-with-holistic-ai, https://www.paconsulting.com/newsroom/hr-dive-how-to-audit-ai-tools-for-bias-18-july-2023) • Develop or Adopt Bias Audit Methodologies and tools (IBM AI Fairness 360, Google What-If Tool, Fairlearn (Microsoft), AI Audit Checklist (AI Now Institute)
Transparency	• Consider adopting open-source practices such as reproducibility, discoverability and sharing (e.g Hugging Face Model Cards,

	Open Source Generative AI Guide – GitHub Copilot Docs) - Use clear markers to label generative AI content e.g this content was generated with AI - Document your AI use case to state how and why AI is used - Explain and share what type of data was used to train or fine tune your pre-trained AI model
User Communicati on	- Have feedback mechanisms and regularly hold feedback sessions with stakeholders and users. - Put strategies in place to show how user related feedback has led to AI related updates and changes
Explainability	- Keep logs of outputs and sometimes publish these logs for user review - Develop a mechanism to allow users request for human review
Organisationa l preparedness	- Develop a documented AI policy guiding the generative AI lifecycle from project initialisation to development and operational phase. - Roles and responsibilities related to AI risk management should be clearly defined across teams. - Staff including developers/technical team should have context specific knowledge of the domain of the AI solution
Financial risks	- Carry out an upfront evaluation of cloud storage costs and have strategies in place

	maintain your storage as your data increases • Be aware of data protection audit costs • Have a clear financial plan for integrating new services into the existing product • Understand the use of tokens and be aware of the upfront cost when tokens are exhausted. Use methods such as summarisation and extraction, contextualisation, continuous conversation with informed responses, and iterative updating • have a customer pricing model targeting various tiers of customers using your generative AI services • explore cost-effective infrastructure solutions for deploying generative AI. e.g open-source models, agent-based cloud platforms, and Hug facing
Technology reliance and Deskilling	• Explore training covering a broad spectrum of capabilities from technical proficiency in AI systems to ethical considerations and impact assessment • ensure that the team is not only adept at using AI tools but also vigilant about their implications and the ethical considerations surrounding their use. • Provide access to affordable training or upskilling programs for existing staff to support AI deployment and maintenance • Develop internal champions who can train others, promoting a multiplier effect.

Shadow use	• Regularly train employees on the risks of Shadow AI and responsible AI usage. • Create and disseminate an organisational AI use policy that defines acceptable and unacceptable AI use, including specific guidance around generative AI • Provide safe, organisation-sanctioned AI tools to fulfil legitimate use cases. • Encourage employees to share new AI use cases transparently and collaborate with IT/governance teams to assess viability.
Talent and Human Resources	• Tap into academic partnerships: Partner with local universities or research centres to access student talent through internships, co-ops, or collaborative projects • Engage with global AI talent on platforms like Hugging Face, GitHub, or AI-specific Slack/Discord groups • Use Low-Code / No-Code Generative AI Tools • Build a Talent Retention Strategy
Data protection	• Adopt the Process, Policies, and Technology (PPT) framework • Under the Process-Oriented Mitigation Strategies consider strategies for Responsible Data Management which include Implementation of clear processes for Data storage, access, deletion, and sharing. Data Protection: Processes to address: Data subject rights. Data protection impact assessments (DPIAs). Identification of lawful basis for data

	processing. Data Quality and Diversity: Ensuring that data used is accurate, representative, and inclusive. ● Development of internal policies and strategies, such as: Data Access Policies: Who can access data and under what conditions. Data Generation and Deletion Policies. Data Sharing Policies, including controls over third-party data usage. Creation of a unified Data Governance Policy Document to consolidate all data-related policies and ensure coherence. ● Deployment of privacy-preserving technologies to safeguard data: End-to-end encryption. Differential privacy. Homomorphic encryption. Federated learning. ● Use of Consent Management Systems to handle user permissions and transparency. ● Integration of Data Auditing and Monitoring Tools to detect misuse, ensure compliance, and maintain accountability.
Intellectual Property	● Assess Dataset Provenance and Licensing Conditions (Use resources like https://spdx.org/licenses/ to decode license types, maintain data sheets or cards using resources such as https://arxiv.org/abs/1803.09010 and https://huggingface.co/docs/datasets/dataset_cards ● Conduct Due Diligence on Third-Party Providers

	• Enter into Contractual Agreements with IP Assurances (Stanford Human-Centered AI Contract Clauses, World Intellectual Property Organization (WIPO) – AI and IP, Mozilla Open Source AI Contracts Guide, UKIPO Guidance on IP and AI • Do Not Rely Solely on Developer Indemnities
Contextual/Local representation	• Source data locally; involve regional content creators and cultural experts in dataset curation • Create simulated dialogues, images, or responses for local norms, holidays, idioms, and taboos • Establish governance bodies that guide development and deployment. By including representatives from communities where the AI will be used • Clearly communicate the cultural limitations and boundaries of the AI system. • Allow users to report culturally inappropriate or inaccurate outputs. Implement in-app feedback tools that log and learn from such reports
Platform Risks	• Diversify Providers (Avoid Single-Point Dependency). If your product uses OpenAI, also test and integrate equivalents from Anthropic, Cohere, or open-source models like LLaMA or Mistral. • Where feasible, run open-source models on-premise or in your own cloud environment for critical workloads.

	- Implement cost monitoring and budgeting tools to detect sudden price spikes in API usage. - For enterprise use, negotiate Service Level Agreements (SLAs) that include uptime guarantees, price change notice periods, and data ownership clauses. Example: Request a 6-month notice period for any pricing structure changes. - Use an abstraction layer in your application so that switching models or providers is technically feasible without a full rebuild. - Conduct Risk Assessments and Exit Strategy Planning
Sector Specific Risks	
Health	- Collaborate with insurers to explore or co-develop bespoke insurance products that cover bodily injury or harm resulting from AI-enabled consultations - Develop APIs and standards that align with your contextual national digital health frameworks - Ensure that all AI-generated outputs are reviewed and verified by licensed medical professionals before being delivered to patients. - Offer digital literacy training to patients and clinicians about how the AI system works, its accuracy, and its limitations. - Seek certification from recognised health tech regulators and conduct regular third-

	party audits to validate clinical safety and performance. • Involve diverse health professionals and patient representatives in system design and testing. • Tune algorithms based on local epidemiological and cultural data • Partner with primary healthcare centres and state health boards to co-deploy the platform in pilot locations. • Join any national AI or health innovation sandboxes to test products
Art	• Implement a voluntary opt-out policy allowing artists to exclude their work from training datasets. • Adopt a neutral stance on copyright ownership of AI-generated art, clarifying that neither the SME nor the users hold exclusive rights. • Develop mechanisms such as word bans and internal moderation tools to limit generation of inappropriate content. • Engage in ongoing discussions with platforms such Google and review app store policies to ensure full compliance in future updates. • Continue partnerships with living artists and institutions to show support for human creativity. • Provide consistent templates and recipes for AI to follow, developed by cultural experts to ensure fidelity

	• Carry out moderation of text prompts using OpenAI's moderation API and an internally developed moderation model to prevent inappropriate input. • Ensure you have a cultural director role in place or set up a culture audit unit
Journalism	• Have a framework embedded in your editorial policy to flag and correct biased news content. • Adapt international AI journalism frameworks to fit your sociocultural context e.g journalismai.info • Establish clear internal rules distinguishing when content requires disclosure of AI use versus when it's considered editorial assistance. • Have a policy in place guiding using generative images to depict real events where no actual footage exists and in non-accessible areas. • Consider reassigning tasks thereby reducing job displacement anxieties by journalists e.g AI handles presentations; human journalists continued investigation and editing. • Seek more localised AI platforms or use alternative tools to improve pronunciation of indigenous words. • Use clear disclosures: Each AI avatar introduces itself as an "artificial

	intelligence presenter" to prevent confusion. • Educate audiences on what AI is and why it is being used. • Establish a media Lab (like a media R&D sandbox) to test new features
Education and Corporate Training	• Use domain experts (e.g., national curriculum, teachers) to train and validate AI-generated content. • Include metadata or confidence scores alongside AI outputs. • Partner with local educators and stakeholders to ensure inclusive development and deployment • Provide low-cost or free access options for underserved communities • Align development with global and regional guidelines •
Legal	• Ensure compliance with Legal professional regulatory bodies (e.g Solicitors Regulation Authority (SRA) in the UK context) • Ensure that client data is never used for training LLMs. • Integrate privacy-by-design principles in system architecture and workflows • Select LLMs that are well-suited and tested for legal domain use. • Monitor outputs regularly with automated evaluations and human oversight by legal engineers.

	• Share objective examples of model performance to reduce uncertainty and build trust. • Use the LLMs’ built-in safety filters as a baseline but not a substitute for internal AI policies and guidelines.
Financial Investment	• Consider establishing a Financial Academy to empower users with the knowledge to understand and question AI-driven advice. • Limit LLM’s Functionality to educational use only — not actual investment advice • Maintain a clear boundary between informative content (e.g., compound interest explanations) and regulated financial advice • Use a chatbot to increase awareness and financial literacy but defer personalised recommendations to licensed channels or tools already approved under regulation • Retain records of advice separately from LLM outputs to comply with advisory regulations. • Use a phased deployment with separate bot functions (onboarding bot, support bot, operational bot). • Consider using generative AI sparingly in high-stress support interactions to avoid user frustration—escalate to human agents when needed
Corporate Intelligence	• Implement and enforce a comprehensive internal GenAI use policy that explicitly prohibits entering proprietary or sensitive company data into public GenAI platforms.

	• Implement systems to identify, tag, and manage PII (Personally Identifiable Information) to avoid unauthorized retention or processing • Prior to embedding GenAI in workflows, verify third-party compliance (e.g., OpenAI) with legal requirements like data erasure and subject access request protocols. • Require that explanations and sources accompany AI-generated profiles or summaries. • Embed ethical review checkpoints during product development, especially when using GenAI in customer-facing analytics • Establish review mechanisms for outputs, especially in domains such as risk evaluation, legal analysis, or client intelligence • Invest in reusable extraction templates for PDFs and OCR tools that integrate with data cleaning pipelines.
Platform Provider (decentralised agent based)	• Establish clear ethical guidelines for node operators on the use of copyrighted data • Implement content usage tracking or watermarking to ensure AI outputs respect original creators • Introduce a reputation system or governance mechanism for node activity and compliance. • Allow for community-driven oversight using transparent, decentralised voting mechanisms

	• Ensure regular audits and peer reviews of data handling methods across nodes • Limit speculative uses of blockchain by using NFTs strictly as software licenses • Maintain jurisdictional flexibility by avoiding deep entrenchment in specific chains (e.g., Ethereum)

3.4.1 Create a generative AI policy.

An essential initial part of the process for mitigating against the risks identified above is to create a policy on the use of generative AI. When implemented effectively, this can lead to both internal and external benefits, including setting clear expectations for employees to abide by, providing certainty and transparency to customers, and supporting regulatory compliance and adherence to ethical AI principles. Whilst the exact content will vary between different SMEs, any organisational generative AI policy should, as a minimum, clearly identify acceptable and responsible uses as well as relevant restrictions on inappropriate uses.[7] In addition, any such policy should account for potential non-compliance by establishing a procedure for reporting, investigating and resolving alleged breaches. The following subsections elaborate further on the key aspects of a generative AI policy for SMEs, the exact features of which will need to be monitored and updated as necessary, starting with the need to form responsible data governance practices.

[7] See, e.g., https://www.gov.uk/government/publications/guidance-to-civil-servants-on-use-of-generative-ai/guidance-to-civil-servants-on-use-of-generative-ai

3.4.2 Ensure responsible data governance practice.

To mitigate ethical, legal, technical, and socio-cultural issues associated with the use of data in AI, SMEs need to develop responsible data governance mechanisms based on Process, Policies and Technology (PPT) framework. Developing appropriate processes through reasoned policies and agile technologies. Some of these processes for data management (e.g., storage, access, deletion, sharing), data protection (e.g., processing for addressing data subject rights, data protection impact assessment, identification of lawful basis for processing), ensuring data quality and diversity. These processes must have foundations in internal policies and strategies. For instance, data access policy that defines who has access to data and conditions for this access, policies on data generation and deletion as well as data sharing policies. These can as well be put into one big policy document called data governance policies. Ways to control data given to 3rd parties that are subsequently used to train AI systems can be included in this policy document. The processes and policies can be tied together with agile technology to help the preservation of privacy and confidentiality, avoidance of data leakage and compliance with legal provisions and ethical principles. These can include privacy preserving technologies (e.g., end-to-end encryption, differential privacy, homomorphic encryption, federated learning), consent management systems, data auditing and monitoring tools.

3.4.3 Ensure appropriate skill mix and training.

In mitigating risks associated with generative AI, it is crucial to focus on the composition of the team's skills and the training they receive. Ensuring a balanced skill set is foundational for the responsible use of generative AI technologies. This involves assembling a team where each member's expertise complements the others, covering a broad spectrum of capabilities from technical proficiency in AI systems to ethical considerations and impact assessment.

The significance of having a well-rounded skill set cannot be overstated. Competencies are needed across various levels of AI engagement, ensuring that the team is not only adept at using AI tools but also vigilant about their implications and the ethical considerations surrounding their use.

To leverage generative AI effectively while minimizing risks and preventing harm, targeted training for employees is indispensable. This training should not only aim at enhancing their technical abilities but also at fostering an understanding of the ethical use of AI. It is about enabling the workforce to harness the potential of generative AI in a way that adds value to your enterprise, while also instituting practices that ensure responsible usage.

Encouraging responsible use among staff involves creating an environment where ethical considerations are at the forefront of AI deployment. By investing in continuous education and promoting a culture of ethical awareness, SMEs can navigate the complexities of generative AI, ensuring that their use of such technologies is both innovative and conscientious.

3.4.4 Select appropriate Large Language Model (LLM)

In selecting an appropriate LLM under responsible generative AI, it is crucial for SMEs to first understand specific business needs and use cases. Evaluating model capabilities such as language understanding, generation quality, and computational requirements, alongside considering ethical and responsible AI principles like bias mitigation, transparency, and accountability. Assessing data privacy and security measures, as well as the model's fairness and inclusivity, ensures protection of sensitive information and promotes ethical AI practices. Scalability, cost considerations, vendor reputation, and support also play key roles in the selection process. Pilot testing and continuous monitoring allow for validation of the LLM's performance and ongoing improvement, while staying informed about evolving standards and best practices ensures ethical and compliant use of generative AI technologies within SMEs. This should include an understanding of the limitations of particular models with respect to use cases and the potential for hallucination. SMEs may also wish to consider the use of open-source models with published hyperparameters, and models trained on open data, enabling knowledge sharing across industry and fostering a culture of transparency in development with benefits to security and user privacy.

3.4.5 Ensure transparency.

In technology integration, transparency is crucial for building trust, meeting ethical standards, and addressing concerns. SMEs integrating generative AI systems need to prioritise AI models that provide clear explanations for their decisions; models that are interpretable or that are inherently transparent. It is also important to maintain comprehensive documentation of the AI design and development processes wherever they are involved; detailing data sources, pre-processing steps, model architectures and parameter tuning. or their decisions; models that are interpretable. It is also important to maintain comprehensive documentation of the AI design and development process which can be made accessible to relevant stakeholders. Ensuring transparency can also include considering open-source practices where applicable and conducting regular audits of the systems’ algorithms and decision-making processes.

3.4.6 Consider unintended impacts.

The use of generative AI can have unintended effects that negatively impact upon a company's reputation. Consider two: bias, and sustainability challenges.

Bias. The value of AI lies in its ability to process vast amounts of data. However, this comes with the risk of perpetuating existing biases. Small enterprises may lack the resources to procure or make independent analyses of large, diverse datasets, leading to AI systems that may not be trained on representative data, or to a dependence upon external processing systems. This can result in AI models that inadvertently discriminate against certain groups, particularly in hiring, lending, and customer service applications. The consequences of AI bias for small businesses are not merely ethical but also practical. Biased AI systems can alienate segments of the customer base, damage brand reputation, and lead to legal challenges. Moreover, these biases can skew market research and business insights, leading to poor strategic decisions. Small enterprises, with their limited buffers against market fluctuations, are particularly vulnerable to such missteps.

Sustainability. Sustainable AI presents a special challenge for small businesses that may wish to claim sustainability credentials. The development and maintenance of AI systems can be costly, and the expertise required to ensure they are sustainable may be beyond the reach of many small businesses. Moreover, the energy consumption associated with running AI algorithms is a significant concern, as it can contribute to a business's carbon footprint. Small enterprises must balance the benefits of AI with the need to minimize environmental impact, which can be a delicate and complex task.

3.4.7 Consider how much you want to reply upon generative AI for business-critical functions.

The involvement of generative AI in business processes creates a potential for increased operational complexity. While AI systems promise to streamline processes, the reality is that they often require significant management and oversight. Small businesses may find themselves grappling with the intricacies of AI integration, from data management to system maintenance, which can be both time-consuming and costly.

Another risk is that of alienating customers. As AI becomes more prevalent in customer service, there is a possibility that the personal touch, which is often a unique selling point for small businesses, will be lost. Customers who value human interaction may feel disconnected from a business that leans too heavily on automated systems, potentially leading to a loss of loyalty and a reduction in revenue.

3.4.8 Consider intellectual property.

For enterprises using or looking to use generative AI, a core area of concern is the potential risk of liability arising from infringement with protected intellectual property (IP) rights, such as copyright.[8] Both initial training and more context specific fine-tuning of generative AI models requires a significant volume of data, a common source for which is those data that can be extracted from websites and other online environments through web-scraping and other data harvesting techniques. Whilst some of these sources are likely to be made available as open access, others may subject to more restrictive licensing conditions for reuse, such as that users request permission

[8] See, e.g., https://www.acrolinx.com/wp-content/uploads/2023/08/Acrolinx_Generative-AI_Report_FNL.pdf

and/or pay a fee. Indeed, even freely *accessible* sources of data may be restricted in this way. As such, SMEs need to actively consider IP in their use of generative AI and take steps to mitigate against the risks.

For SMEs involved in training or fine-tuning of generative AI models, whether developed internally or, perhaps more likely, where provided by a developer via an application programming interface (API), the provenance of the dataset(s) and any conditions (e.g., licensing and compensation) attached to usage should be assessed to promote transparency (see above) and ensure full compliance with IP law. If the data used and/or the generative AI model itself is sourced from a developer or another provider, SMEs will need to ensure they carry out sufficient due diligence. In particular, consideration should be given to entering into contractual arrangements on the basis of assurances around compatibility with IP law and protection against losses from a failure to do so. Some developers, especially larger organisations, are already offering indemnities against legal and financial losses. Whilst also facing legal action from the New York Times for alleged copyright infringement, Open AI, for instance, has established a "Copyright Shield" for all "generally available features of ChatGPT Enterprise", through which it pledges to "pay the costs incurred" for customers facing legal action over copyright infringement.[9] However, as in this example, these indemnity clauses are unlikely to be broadly applicable and, as such, may offer only limited protection for SMEs against the risk of third-party IP infringement arising from the use of generative AI.

[9] https://openai.com/blog/new-models-and-developer-products-announced-at-devday

3.4.9 Make the best use of support from government and other sources.

There are some sources of support available to SMEs that may be able to help them explore or implement generative AI. This can include financial support from governments, funding bodies, universities, but also networks and industry bodies.

- The UK government provides tax relief on corporation tax for Research and Development (R&D) projects, allowing companies to claim back some of the costs of innovative projects in science and technology. Not all uses of generative AI will qualify, but if a project does qualify, then this can be a way of reducing the costs of responsible generative AI deployment. To qualify, a project needs to look for an advance in a field related to your company's trade, have overcome a scientific or technological uncertainty, and do something that could not be easily worked out by a professional in the field. More guidance is available from HMRC.[10]
- The Innovate UK Bridge AI programme is intended to support UK business through the responsible adoption of AI in priority sectors, such as transport, construction, agriculture and food production, and creative industries. The programme offers funding and support to help innovators assess and implement trusted AI solutions, connect with AI experts, and elevate their AI leadership skills.[11] Support from the programme includes:

[10] https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief

[11] https://iuk.ktn-uk.org/programme/bridgeai/ open support opportunities are advertised here: https://iuk.ktn-uk.org/opportunities/?_sft_areas=bridgeai

  - Access to the Hartree Centre Training portal of training and upskilling in AI.[12]
  - Free training courses from the Turing institute, developed by Trilateral Research on operationalising ethics in AI[13]
  - This included a now closed funding competition for UK organisations to support the development and adoption of AI and machine learning solutions[14]
- There are three regional SME engagement hubs in Cardiff, North East England and Northern Ireland providing local support from organisations with expertise in data science and AI to help SMEs.[15]
- Schemes like the AI Futures Grant provide support to SMEs to meet the relocation costs of early to mid-career AI researchers and engineers with exceptional promise moving to the UK.[16]

3.4.10 Future Development Opportunities

- The SEO market is likely to be disrupted in the near future.[17] As SMEs navigate this evolving landscape, it is crucial to monitor SEO metrics closely to gauge any significant shifts

[12] https://iuk.ktn-uk.org/opportunities/bridgeai-access-to-hartree-centre-training-portal/

[13] https://www.turing.ac.uk/courses/operationalising-ethics-ai-intermediate

[14] https://apply-for-innovation-funding.service.gov.uk/competition/1714/overview/c85b3edc-0416-4e92-a5d3-014fe8aab148#eligibility

[15] https://www.hartree.stfc.ac.uk/sme-hubs/

[16] https://www.gov.uk/guidance/corporation-tax-research-and-development-rd-relief

[17] https://www.forbes.com/sites/nishatalagala/2023/03/25/chatgpt-is-seoas-we-know-itdead/?sh=12d6caaa395e

resulting from the increasing prevalence of Generative AI. Ongoing observation and adaptation will be key to staying ahead in such a dynamic environment. For example, as the popularity of chat-based interfaces continues to rise, it's essential for SMEs to reassess their search marketing strategies and how they align with these evolving trends. For instance, if your approach has centred on creating content-rich pages, like blogs, to attract visitors through search queries, you may need to consider the potential impact of chatbots autonomously providing answers to user inquiries. Conversely, if your objective is to showcase specific offerings, such as upcoming events or classes, the influence of chatbots might be less pronounced.

- Experimentation and continual measurement - to exploit future development opportunities while mitigating risks SMEs can adopt four different layers of potential mitigations: at the **model level**, at the **safety system level**, at the **application level**, and at the **user level.** Adopting this strategy will enhance an iterative and layered approach to identify and mitigate risks while maximising future development opportunities tailored towards individual needs and staying ahead in a dynamic generative AI adoption environment.
- As Generative AI has the potential to revolutionize content creation by generating high-quality images, videos, and graphics, and by creating dynamic content tailored to an individual’s preferences. This can help businesses stand out in a crowded marketplace and optimize engagement and conversion rates.

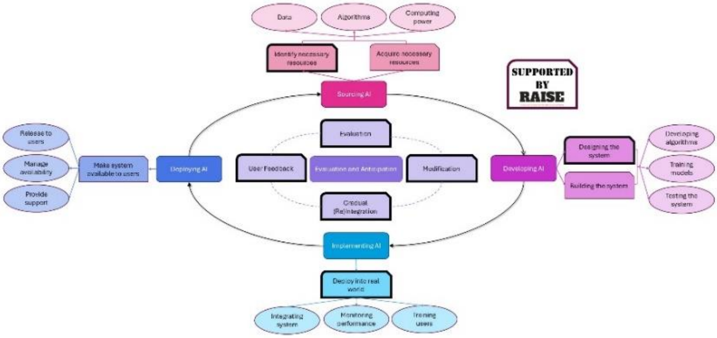
In 2023, a New Zealand-based supermarket Pak 'n' Save created a meal planner app (Savey meal-bot) designed to help households save money and reduce food waste by suggesting meals complete with a full recipe and directions. A user shared on X (formerly Twitter) *how this app suggested a recipe for deadly chlorine gas.*

3.4.11 Ensure Human Oversight

The concept of 'human-in-the-loop' (HITL) is a term in Innovation that refers to the continuous involvement of human oversight in the design, development, and deployment of AI systems. Whilst acknowledging that SMEs often do not have sufficient resources for dedicated AI experts on the team, it is critical that human oversight ensure that people with knowledge can monitor relevant stages of design and implementation. HITL can help the SME address ethical considerations, comply with regulations, become transparent and proactively detect errors. This can be achieved through;

- A robust review and approval of all recommendations made by AI systems before implementation. The team's judgement and experiences are thus leveraged on to ensure that final decisions align with business goals and ethical considerations.
- Identification and handling of complex or unexpected situations that AI systems are not built to address.
- Clear establishment of roles and responsibilities within the organisation regarding who makes the final decision between the AI and humans. It is also essential for SMEs to maintain open communication with human experts and AI designers, providing feedback on performance for improvement of the system.

5 AI Lifecycle Policies and Approaches



Generative AI lifecycle approach

As a fundamental aspect of the Responsible Generative AI for SMEs guidelines, our aim is to seamlessly incorporate our recommendations throughout the various stages of AI adoption by SMEs. This entails integrating responsible generative AI principles, recommendations, and practices within the following stages of the AI adoption lifecycle:

5.1 Lifecycle

The AI Lifecycle encompasses the stages in which an AI solution is conceived, developed, deployed and utilized.

- **Sourcing AI** is the first stage of the AI lifecycle and involves identifying and acquiring the necessary resources to develop and implement an AI system. This includes identifying the appropriate data, algorithms, and computing power. **Responsible AI** can be integrated into this phase by:

  - Ensuring that the data used to train the AI system is **high-quality, unbiased, and representative**.
  - Ensuring that the algorithms used to develop the AI system are **transparent, explainable, and fair**.
  - Ensuring that the computing power used to develop and implement the AI system is **sustainable**.

- **Developing AI** is the second stage of the AI lifecycle and involves designing and building the AI system. This includes developing the algorithms, training the model, and testing the system.

  Responsible AI can be integrated into this phase by:
  - Ensuring that the algorithms are designed to **meet the specific needs** of the organization.
  - Ensuring that the model is **trained on a representative dataset.**
  - Ensuring that the system is **tested thoroughly to identify and mitigate potential risks.**

- **Implementing AI** is the third stage of the AI lifecycle and involves deploying the AI system into the real world. This includes integrating the system into existing processes, training users on how to use the system, and monitoring the system's performance.

  Responsible AI can be integrated into this phase by:
  - Ensuring that the system is **implemented in a way that minimizes risks**.
  - Ensuring that the system is **monitored closely to identify and mitigate any unintended consequences**.
  - Providing **training for users on how to use the system responsibly**.

- **Deploying AI** is the fourth stage of the AI lifecycle and involves making the AI system available to users. This includes releasing the system to the public or specific user, managing the system's availability, and providing support to users. Responsible AI can be integrated into this phase by:

  - Ensuring that the system is **deployed in a way that minimizes risks**.
  - Ensuring that the system is **monitored closely to identify and mitigate any unintended consequences**.
  - Providing **training for users on how to use the system responsibly**.
  - Ensuring that the system is **designed to be fair and equitable**.
  - Ensuring that the system is **transparent and accountable**.

- **Continuous Evaluation and Anticipation**: This focuses on a longer approach plan and involves continuous evaluation and modification/gradual integration of new services while constantly obtaining user feedback for systems improvement. This phase also focuses on anticipating future requirements and adaptation.

We believe that integrating our responsible guidelines and leveraging the AI lifecycle and approaches across sourcing, developing, implementing, and deploying stages can help SMEs harness the benefits of AI while mitigating risks and maximizing return on investment.

6 Next Steps

The RAISE Guidelines provide guidance on the responsible use of generative AI for small and medium-sized enterprises (SMEs) in the UK and Africa. These guidelines aim to gather feedback on content and representation, as well as to assist SMEs in avoiding potential pitfalls associated with generative AI use.

Key findings of the guidelines highlight both the risks and mitigation strategies associated with generative AI. Risks include platform risks, negative impacts on employee job satisfaction and feelings of job security, and "shadow AI." Mitigation strategies include establishing a generative AI policy, ensuring sound data governance practices, and acknowledging the additional effort and time required by SMEs to implement responsible AI practices. The guidelines also outline four key steps to integrate responsible AI into the AI lifecycle: sourcing AI, developing AI, implementing AI, and deploying AI. In the deployment phase, responsible AI involves minimizing risks, monitoring the system for unintended consequences, providing user training, ensuring fairness and equity, and guaranteeing transparency and accountability.

7 Appendix A: Frameworks and principles for responsible AI

Currently, several global guidelines reflecting various principles for responsible AI adoption exist. These include:

- Responsible AI Adoption: 5 Steps Every Startup Must Take[18]
- OECD AI Principles[19]
- Trustworthy AI principles[20]

[18] https://yourstory.com/2023/11/startup-guide-ethical-ai-practices

[19] https://oecd.ai/en/ai-principles

[20] https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

- EU AI Act[21]
- UK Algorithmic Transparency Recording Standard[22]
- Edinburgh declaration on responsibility for responsible AI[23]
- UNESCO Guidance for generative AI in education and research[24]

In the African landscape, steps have also been taken to recommend guidelines containing principles that recognise particular and distinctive socio-cultural nuances. Some of these initiatives include:

- AI For Africa Blueprint[25]
- African Union Resolution 473 on human and people's rights and AI[26]

The guidelines above reflect various AI principles as shown in table 1 below.

Principle	Description

[21] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=celex%3A52021PC0206

[22] https://www.gov.uk/government/publications/algorithmic-transparency-template

[23] https://medium.com/@svallor_10030/edinburgh-declaration-on-responsibility-for-responsible-ai-1a98ed2e328b

[24] https://unesdoc.unesco.org/ark:/48223/pf0000386693

[25] https://smartafrica.org/wp-content/uploads/2023/11/70029-eng_ai-for-africa-blueprint-min.pdf

[26] https://achpr.au.int/en/adopted-resolutions/473-resolution-need-undertake-study-human-and-peoples-rights-and-art

Inclusive growth, sustainable development and well-being[27]	This principle emphasizes the importance of guiding AI development and use toward prosperity and beneficial outcomes for both people and the planet, highlighting its role in advancing inclusive growth, sustainable development, and well-being. It acknowledges the risks of exacerbating inequalities and biases, advocating for responsible stewardship, multidisciplinary collaboration, and public dialogue to ensure AI empowers all members of society and fosters public trust.
Human-centred values and fairness[28]	AI development should adhere to human-centred values, including fundamental freedoms, equality, fairness, and privacy, to ensure that it aligns with human rights and promotes social justice. Implementing safeguards, human intervention, and oversight in AI systems can protect human rights, reduce discrimination, and build public trust, while measures like human rights impact assessments and ethical codes further promote fairness and value alignment.
Transparency and explainability	This principle is about transparency and responsible disclosure around AI systems to ensure that people understand when they are engaging with them and can challenge outcomes.
Robustness, security and safety	AI systems must function in a robust, secure and safe way throughout their lifetimes, and

[27] https://oecd.ai/en/dashboards/ai-principles/P5
[28] https://oecd.ai/en/dashboards/ai-principles/P6

	potential risks should be continually assessed and managed
Accountability	Organisations using AI should be accountable for the proper functioning of AI systems and for the respect of the other principles, based on their roles, the context, and consistent with the state of art.
Decolonisation	Decolonisation of AI involves challenging colonial perspectives and creating more inclusive, culturally varied approaches to artificial intelligence. It aims to undo colonial legacies and envision alternative futures beyond dominant AI paradigms.
Solidarity	Solidarity as an AI principle shares the prosperity created by AI, implementing mechanisms to redistribute the augmentation of productivity for all; and shares the burdens, making sure that AI does not increase inequality and no human is left behind. Solidarity as an AI principle also assesses the long-term implications before developing and deploying AI systems, so no groups of humans become irrelevant because of AI systems

8 Appendix B: Existing guidance and support

8.1 Organisations

- Turing Institute
- Ada Lovelace
- Journalismai

8.2 Training

- BRIDGE AI
- https://www.holisticai.com/blog/technical-resources-bias-mitigation
- https://www.ibm.com/opensource/open/projects/ai-fairness-360/

8.3 Repositories

- https://aistandardshub.org/
- https://interpret.ml/