KPMG

From Promise to Practice

The Risks and Governance Approach of Generative AI

—

July 2023

Executive Summary

Generative Artificial Intelligence (AI) models such as OpenAI's ChatGPT and Google's Bard are poised to revolutionise the way businesses operate by enabling them to automate and execute a range of tasks with unparalleled speed and efficiency.

These models are able to generate outputs that resemble human-created content, such as text or other media formats (image queries in the latest GPT-4 model) based on a user prompt. Many people believe that this technological progress marks the beginning of a new era for Artificial Intelligence and has the potential to lead the way for the fifth industrial revolution. Since its development, a wide range of applications based on Generative AI been created such as assistants for search engines, tools to enhance office productivity, helpers for coding, virtual tutors for education, and many others.

However, like a double-edged sword, training and using General AI models have also sparked significant worries regarding ethical and legal matters, as existing regulatory frameworks struggle to keep up.

This paper explores regulatory realm, legal risks, and ethical considerations associated with the adoption of Generative AI and proposes a recommended governance approach to address these challenges.

By implementing a robust governance framework, organisations can mitigate legal and ethical risks while maximising the benefits of Generative AI. This approach strikes a balance that promotes innovation, upholds fundamental values, and protects the rights and well-being of individuals and society as a whole.

Data Privacy

The Generative AI models were trained on vast amounts of data including personal or sensitive information, raising concerns regarding compliance with GDPR regulations.

Intellectual Property

The Generative AI models were trained on a large body of created works and the model may produce content that closely resembles existing protected works, raising legal concerns over potential copyright infringement.

Factual Accuracy

The risk of generating false information, i.e., hallucinations, is a major challenge with Generative AI models, especially when outputting large amounts of information or operate in domains where they lack adequate training, such as in a medical setting.

Discrimination

Generative AI can raise civil rights issues if the generated content discriminates against protected groups.

Liability

Under the EU Product Liability Directive and proposed AI Liability Directive, the operator may be held liable for harm caused by an AI system to third parties.

Regulatory Realm

🏛 **In the regulatory realm, there is an ongoing effort to establish guidelines and strengthen the existing frameworks, including the possibility of enacting laws. The primary objective of these endeavours is to ensure the protection of individuals from the potential risks posed by the technology.**

It is crucial to build citizens' trust in the development of AI and this can be achieved with landmark legislation that must resist the challenge of time. Following this line of reason, the European Parliament has voted by a large majority in favour of adopting a wide-ranging proposed law on AI, which is the first relating to AI by a major regulator.[1]

The proposed law classifies applications of AI to three risk categories. First, applications and systems that intentionally employ manipulative techniques or engage in social scoring were both given as examples of applications that create an Unacceptable Risk. The second category pertains to High-Risk applications subject to specific legal requirements, such as CV-scanning tools that rank job applicants. Lastly, AI systems such as chatbots fall into Limited Risk category Their compliance obligations are light touch and focus on transparency. Additionally, Generative AI systems, like ChatGPT, would have to comply with transparency requirements: disclosing that the content was AI-generated, ensuring safeguards against generating illegal content and publishing detailed summaries of the copyrighted data used for model training.[2]

The proposed law contains various exceptions. For instance, while "real-time" facial recognition by the police is generally prohibited, there are exceptions if the images have already been captured and the comparison and identification occur only after a significant delay ("post" biometric identification system) or if the technology is employed to locate missing children.

In the UK, the joint discussion paper on AI (DP5/22)[3] by Bank of England (BoE), Prudential Regulation Authority (PRA), and Financial Conduct Authority (FCA) examined the benefits, risks and harms related to the use of AI in financial services and how existing legal requirements and guidance apply to the use of AI, focusing on the aspects from consumer protection, competition, data, model risk management, governance, operational resilience, outsourcing, and third-party risk management. Additionally, the supervisory authorities encouraged all relevant stakeholders to respond to this DP and explore whether the existing sectorial legal requirements and guidance were sufficient to address the risks and harms associated with AI.

In this article, we share our views on the potential risks associated with Generative AI and their implications for organisations. We explore how we can implement or enhance the governance approach to ensure the responsible and beneficial use of Generative AI, aiming to harness technology for the greater good.

Note: (1) European lawmakers vote overwhelmingly in favor of adopting the first AI Act…, https://www.businessinsider.com/ai-act-european-parliament-voted-in-favor-2023-5
         (2) EU AI Act: first regulation on artificial intelligence, https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence
         (3) DP5/22 - Artificial Intelligence and Machine Learning, https://www.bankofengland.co.uk/prudential-regulation/publication/2022/october/artificial-intelligence

Expanding on the Legal Risks

Data Privacy

The Generative AI models were usually trained on vast amounts of data including personal or sensitive information. Taking ChatGPT as an example, according to GPT-4 Technical Report[4], the ChatGPT model was trained on a variety of publicly available and licensed data sources, which may include publicly available personal information. Under the EU General Data Protection Regulation (GDPR) Article 5, personal data collection and processing shall be limited to only what is necessary to fulfil the original purpose (known as "Data Minimisation"). It is unclear if ChatGPT is complying with this rule regarding data collection and processing given it gathered up billions or trillions of data points from Internet.

In addition, under Article 17, people have the right to request their personal data to be erased under certain circumstances (known as "Right to erasure"). It is unclear if ChatGPT or Generative AI models in general supports this due to the nature of the data being processed by these systems. Once a data point has been used in model training, its information will be learned by underlying neural networks and embedded in some of the billions of weights in these networks. It is a very difficult if not impossible task to ask such a large model to completely forget that information. Italy became the first Western country to ban ChatGPT on 31 March 2023 over the privacy concerns relating to this model[5].

Organisations' proprietary data may contain sensitive information that needs to be protected from unauthorised access or disclosure. When you send proprietary data to a third-party service hosting a Generative AI model, you may be giving up some of your ownership rights to that data. Depending on the terms of service of the service provider, your data may be leaked to other third-parties, which may cause a data breach. The GDPR requires that all personal data must be either stored in the EU, or within a jurisdiction that has similar levels of protection. Sending personal data to an non-compliant data centre outside of EU will break the GDPR law.

Intellectual property (IP) risks associated with using Generative AI can arise in several ways.

First, due to the enormous amounts of data that AI systems have been trained on; such training data will likely include third-party IP, such as patents, trademarks, or copyrights, for which use authorisation has not been obtained in advance. As a consequence, outputs from the AI systems may infringe others' IP rights. This phenomenon has already led to litigation[6],[7].

Intellectual Property

Second, disputes may arise over who owns the IP generated by an AI system, particularly if multiple parties contribute to its developments. For example, OpenAI's Terms of Use assign the "right, title and interest" in the output of the LLM to the user who provided the prompts, so long as the user abided by OpenAI's terms and the law. OpenAI reserves the right to use both the user's input and the AI-generated output "to provide and maintain the Services, comply with applicable law, and enforce our policies"[8].

Third, there is the issue of whether IP generated by AI is even protected because, in some cases, there is arguably no human "author" or "inventor." Litigants are already contesting the applicability of existing IP laws to these new technologies.

As the law surrounding the use of AI develops, companies seeking to use Generative AI tools to develop their products should document the extent of such use and work with IP counsels (e.g., the Intellectual Property Office in the UK) to ensure adequate IP protection for their products.

Note: (4)   GPT-4 technical report, https://arxiv.org/abs/2303.08774
      (5)   ChatGPT banned in Italy over privacy concerns, https://www.bbc.co.uk/news/technology-65139406
      (6)   Microsoft want court to toss lawsuit…, https://www.reuters.com/legal/litigation/openai-microsoft-want-court-toss-lawsuit-accusing-them-abusing-open-source-code-2023-01-27
      (7)   Lawsuits accuse AI content creators of misusing copyrighted work, https://www.reuters.com/legal/transactional/lawsuits-accuse-ai-content-creators-misusing-copyrighted-work-2023-01-17/
      (8)   OpenAI terms of use, https://openai.com/policies/terms-of-use

Expanding on the Legal Risks

Factual Accuracy

One of the biggest challenges of using ChatGPT and other Generative AI models is the risk of hallucinations, where the models may generate information that appears to be true but is actually false. For example, when asking ChatGPT (May 24 version) "Who was the first woman accountant to be admitted to the ICAEW by examination in 1924?", ChatGPT answered "Mary Harris Smith". However, the answer should be "Ethel Watts". Whereas Harris Smith became the world's first female chartered accountant, she had not passed the organisation's examinations. This risk increases when the models are asked to generate larger amounts of information or in areas that they were not well trained. It is crucial for users and organisations to ensure that they use factually accurate information for decision making. Failure to do so could result in exposing the company to liability if others rely on the information and suffer harm. For example, a Generative AI chatbot in a medical setting may generate inaccurate information, and advise a patient to take a medication that they are severely allergic to.

Discrimination

The ineffective use of Generative AI technologies can affect the values on which organisations and society are founded could lead to breaches of fundamental rights that might result in discriminating, some with serious legal consequences. Some of these rights include non-discrimination based on sex, racial or ethnic origin, religion or belief, disability, age or sexual orientation[9], protection of personal data and private life, or the right to an effective judicial remedy and a fair trial, as well as consumer protection.

These risks might result from flaws in the overall design of AI systems (lack of human oversight) or from the use of data without correcting possible bias, e.g., the system is trained using only or mainly data from men leading to suboptimal results in relation to women[10].

Since AI models are built by people and learn from data created by people, human predisposition can be weaved into an AI's design, development and implementation. Due to this, users could purposefully manipulate AI systems and chatbots to produce unflattering or prejudiced outputs. For instance, ChatGPT, like other LLMs, can learn to express the biases of the data used to train them. As OpenAI acknowledges, ChatGPT "may occasionally produce harmful instructions or biased content"[11].

Another risk stems from the amplification of existing biases. Generative AI models possess the potential to exacerbate pre-existing biases that pervade society. If the input data provided to the model contains biased or discriminatory content, the generated outputs may reinforce and even magnify those biases. This can further perpetuate discrimination and exclusion, exacerbating social inequalities.

Designers, developers, and deployers of automated systems should take proactive and continuous measures to protect individuals and communities from algorithmic discrimination and to use and design systems in an equitable way.

Note: (9)   Equality Act 2010,
               https://www.legislation.gov.uk/ukpga/2010/15/contents
         (10) Women's health outcomes: Is there a gender gap?,
               https://lordslibrary.parliament.uk/womens-health-outcomes-is-there-a-gender-gap/
         (11) ChatGPT general FAQ,
               https://help.openai.com/en/articles/6783457-chatgpt-general-faq

Expanding on the Legal Risks

Liability

Organisations should be mindful of the potential liability risks associated with the employment of Generative AI technologies and take necessary steps to mitigate them. All the discussed issues related to data privacy, infringement upon intellectual property rights or discrimination against individuals based on protected characteristics, may lead to liability risks, which in turn can lead to legal actions, reputational damage, and financial penalties for organisations or individuals working for them.

Therefore, they should conduct thorough due diligence to understand and address the legal implications associated with Generative AI use. For instance, the terms of using ChatGPT state that the total liability to OpenAI for any claims related to its service will not exceed the greater of the amount paid by the user for the specific service that gave rise to the claim in the 12 months before the claim arose, or one hundred dollars ($100)[8]. So OpenAI essentially holds users fully liable for the use they make of the output.

Even more, Generative AI (and AI in general) technologies may present new safety risks for users when they are embedded in products and services. A lack of clear safety provisions tackling these risks may, in addition to risks for the individuals concerned, create legal uncertainty regarding liability for businesses that are marketing their products involving AI. Market surveillance and enforcement authorities may find themselves in a situation where they are unclear as to whether they can intervene, because they may not be empowered to act and/or don't have the appropriate technical capabilities for inspecting systems.

Organisations that intend to use or plan to use the such AI technologies should consult with legal and insurance professionals to determine the appropriate approach to managing liability, e.g., secure liability insurance.

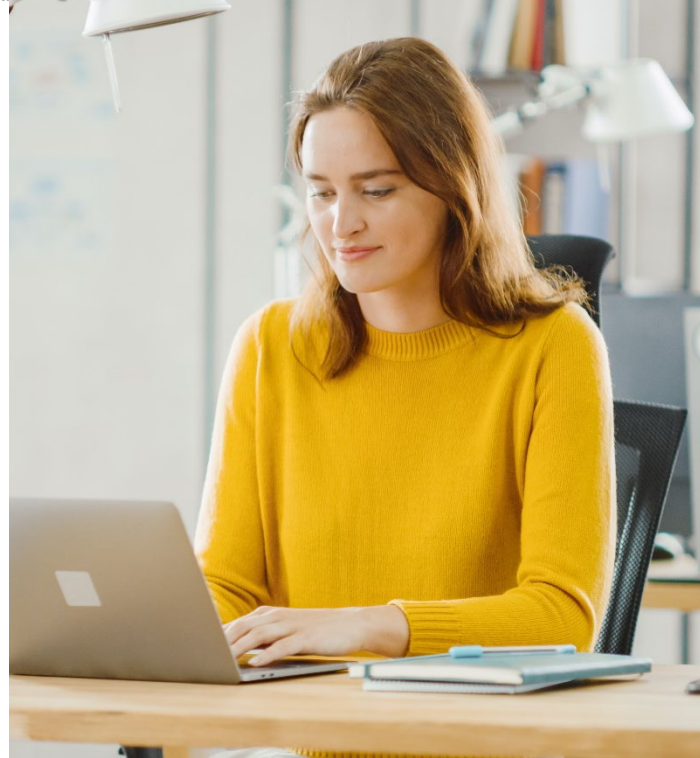Note: (9)   OpenAI terms of use, https://openai.com/policies/terms-of-use

Ethical Risks

> **Like other technologies, there is always a risk of malicious use of Generative AI by individuals or organisations with harmful intent.**

Malicious Use

01

Spreading Misinformation

With the powerful content generation capability of these technologies, drafting a new false narrative can be done at dramatic scale. Researchers predict[12] that propaganda campaigns that we see today such as mass-messaging on social media platforms or long-form news generation on unattributable websites can be easily scaled up to reach a wider audience with much reduced cost.

02

Generating Malware

Model providers have been improving their models to restrict the harmful content from being generated. It is more and more difficult to get the response regarding malicious or abusive activities directly. However, researchers[13] show that they had managed to get the code piece from ChatGPT which could be modified for malicious use by avoiding using malicious keywords in the request. This may lower the technical bar required for creating malware.

03

Targeted Phishing Scams

It may be used by malicious actors to generate phishing messages, designed to trick people into giving away sensitive information or credentials. These messages could be easily customised to target specific people especially vulnerable groups by using Generative AI models.

Unintended Harm

The use of Generative AI may raise the risk of exposure to harmful content leading to unintended harm. While these generative models may have filters in place to prevent explicit or inappropriate language from being used, it's still possible for users to bypass these filters and to come across improper content accidentally[14].

Interactions with AI models may not provide the same level of empathy and understanding as human interactions. Relying solely on AI-generated responses for emotional support may lead to feelings of loneliness or isolation.

Finally, there are debates about the use of Generative AI in education, with some arguing that it could enhance learning while others worry that it could pose a threat and have negative effects on students abilities[15].

Note: (12) Generative language models and automated influence operations: emerging threats and potential mitigations, https://arxiv.org/abs/2301.04246
(13) GPT-4 can't stop helping hackers make cybercriminal tools, https://www.forbes.com/sites/thomasbrewster/2023/03/16/gpt-4-could-help-stupid-hackers-become-good-cybercriminals/
(14) ChatGPT has a sexual harassment problem, https://blog.practicalethics.ox.ac.uk/2023/04/chatgpt-has-a-sexual-harassment-problem/
(15) Open AI in education, the responsible and ethical use of ChatGPT towards lifelong learning, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4354422

Ethical Concerns

Job Security

According to a recent report published by OpenAI[16], their findings reveal that roughly 80% of the US workforce could experience at least 10% of their work tasks affected by LLMs while approximately 19% of workers may witness at least half of their tasks impacted by those models. While the adoption timeline was not given, as technology is evolving, the impacts of LLMs on economics will likely increase, posing rising concerns about job security. Jobs associated with more information processing or repetitive tasks are more likely to be affected such as junior developers, customer service, legal advisors, tax advisors, accountants and auditors, writers, Interpreters and translators.

Transparency and Explainability

Generative AI models possess an inherent "Black Box" nature, which makes it challenging (if not impossible) to understand and explain its internal reasoning and content generation mechanism. This issue has been exacerbated by major AI companies increasingly refraining from disclosing technical details including the training data information in model training process.

Without transparency and explainability, there is little ways of knowing what bias might be included in the system and if the outputs from the models can be trusted. This is especially important when Generative AI models are used for business decision making process. For example, if a company uses a Generative AI model to screen job applications or automate the hiring process, lack of transparency can raise concerns about potential bias. If the model is trained on biased or unrepresentative data, it may inadvertently discriminate against certain demographics, leading to unfair hiring practices.

Plagiarism

Many students are using Generative AI to assist in writing their thesis reports and assignments[17]. Universities and colleges are split over what policies they should put in place. Some have banned the technology over plagiarism fears, while others have chosen to embrace it by asking the students to acknowledge the use of Generative AI models in their works. Technically, it is very difficult to accurately tell the content generated by AI and the content written by human, even with the help of another AI tool. Students may also use the content generated by AI as a basis and subsequently refine and polish it accordingly, which blurs the line.

Plagiarism in the workplace can have far-reaching consequences that can impact organisations and stakeholders. It can create, and not limited to: breaches of trust, reputational damage, negative impact on the productivity (can lead to conflicts and disputes among colleagues), damage to intellectual property, legal issues. It is important for organisations to establish clear policies and guidelines on plagiarism, provide training and education to employees, and take appropriate action when plagiarism is identified.

Environmental Impact

The increasing demand for computing power and data storage required to support large language models like ChatGPT, can have a significant environmental impact. According to a Bloomberg article[18], training of a single AI model can consume more electricity than 100 US homes use in an entire year. While some AI companies state that they have been making effort to improve efficiency and their carbon footprint to run large language models, there are still raising concerns in public about the environmental impact of training and using these models. Organisations need to be aware that the carbon emission generated from these activities will factor into their Scope 2 emission (if use own data centre) or their Scope 3 emission (if use third-party cloud) and the resulting public perceptions.

Note: (16) GPTs are GPTs: an early look at the labor market impact potential of large language models, https://arxiv.org/abs/2303.10130

(17) Universities warn against using ChatGPT for assignment, https://www.bbc.co.uk/news/uk-england-bristol-64785020

(18) Artificial Intelligence Is booming—so is its carbon footprint, https://www.bloomberg.com/news/articles/2023-03-09/how-much-energy-do-ai-and-chatgpt-use-no-one-knows-for-sure

Requirements of AI Adoption and Governance

The number of organisations that turn to AI-power solutions for their business needs is growing exponentially. The new applications of Generative AI offer opportunities for increasing economic and operational efficiency, but they also generate unexpected and unintended consequences and pose new forms of risks that need to be addressed. To enhance the benefits from AI while minimising the adverse risks, organisations need to understand better the scope and depth of the risks posed and develop governance processes and structures to address these challenges.

While there is considerable literature emerging on various aspects of AI, governance of AI is an emerging but significantly underdeveloped area. There is a need to reassess the efficacy of traditional governance approaches, which may be insufficient due to the lack of information and constant changes, and the speed and scale of adoption of AI threatens to outpace the regulatory responses to address the concerns raised.

An effective adoption approach for Generative AI technologies involves thoughtful consideration of the organisation's needs and goals, careful planning, as well as the resources required for successful implementation. This may include selecting the appropriate level of customisation, integrating the model with existing systems and processes, and providing training to employees who will be interacting with the Generative AI systems. Additionally, it is important to establish clear goals and metrics for measuring the success of the adoption, as well as a plan for ongoing monitoring and evaluation.

In terms of governance, organisations must take steps to ensure that the use of ChatGPT aligns with their ethical and legal responsibilities.

Why is AI Governance Important?

The aim of AI governance is to ensure the safety of the public without impeding the technological advancement of AI as well as contribute to the oversight of an organisation's AI tools and supervision. For organisations that want to succeed in the next decades, they must adapt to the evolution of technology and to determine what policies or technical guardrails, should be imposed on Generative AI use in the workplace.

The incorrect adoption of AI applications can lead to undesired consequences. For instance, applications may unfairly discriminate leading to unethical outcomes and possibly reputational damage or may not contribute to the business value of the organisation. Appropriate governance of such applications mitigates the associated risks (fairness, integrity). Moreover, AI governance leads to technological insight (explainability), holds the appropriate persons responsible (accountability), and can increase longevity and security of the application (resilience). The best course of action to mitigate most, if not all, of the risks above is to set up a proper governance framework.

Typically, AI governance models touch on particular aspects, such as fairness or transparency, and focus on specific stages of system development, such as system design. However, organisations need to govern AI systems over their entire life cycles and consider the complete set of requirements in relation to ethics, legislation, and stakeholders.

Governance Approach for Organisations

Our view on the framework for AI governance is that it translates the objectives of responsible AI into nine concrete themes from which organisations can take steps towards better governance and can be grouped into a three-layer framework:
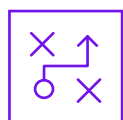
A technical layer

That covers the operational governance for development, management, and use of AI systems, including the AI system and data-driven processes.

A risk layer that covers the

Requirements from the contextual environment looking at the risks and concerns that could arise, and how can these be mitigated.

An organisational layer

That covers the organisational practices, capabilities and the implications of adopting AI for use in the workplace.

Strategic Alignment	**Organisational Alignment**	**User Education & Adoption**
Ethics	**Data & AI Risk Management**	**Regulatory Compliance**
Data & Data-Driven Practices	**Technology of AI System**	**Algorithms**

01 Organisational Layer

The themes stem from the nature of AI governance that would have to ensures the alignment of an organisation's use of AI technologies with its organisational strategies and principles of ethical AI, and educate its adopters.

02 Risk Layer

These themes cover the requirements that organisations should consider in order to mitigate Data and AI risks, i.e., normative regulation, principles and guidelines (self-regulation), and stakeholder pressure.

03 Technical Layer

The themes describe what is necessary to manage the life cycle of algorithm/AI applications and safeguard the quality of the data and technology used.

Strategic Alignment

The strategic alignment requires the definition of an organisational strategy for the use of AI. This provides a general direction and manages expectations regarding the overall set of AI systems that the organisation intends to use and what they are meant to achieve. The strategically aligned use of AI requires organisational resources, capabilities, and processes that can be ensured through management commitment and staff training. Firms have already or are in the process of updating their policies, guidelines, principles and procedures for ethical use of AI. Generative AI is a specific case and may require additional review and uplift to existing frameworks.

Organisational Alignment

The use of Generative AI involves the challenges from several multidisciplinary area and, to this aim organisations should establish a multidisciplinary governance team with expertise in areas such as AI, legal, ethics, privacy, change management (i.e., governance committee) to oversee the use of Generative AI in the workplace.

This structural governance should have mechanisms in place to define reporting structures, governance bodies, and accountability. They comprise roles and responsibilities and the allocation of decision-making authority.

Governance Approach for Organisations

User Education & Adoption

A key step for user adoption is to provide appropriate training and education to employees who will be working with Generative AI to ensure that they understand the technology and its potential risks, benefits and limitations. Timely communications should be in palace to communicate the potential work pattern change or impact to the stakeholders.

Organisations take great care in creating design recommendation for explaining Generative AI, and AI in general, to end users. These recommendations can be categorised into four groups:

- **General:** Recognise that there is no one-size-fits-all type of solution.

- **Addressing the "when":** provide explanations on demand, not all the time.

- **Addressing the "what":** ensure that users understand the underlying factors and reasoning.

- **Addressing the "how":** personalise explanations, link explanations to users' mental models and ensure visibility and discoverability of explanations.

Ethics

Like other forms of AI, Generative AI can influence a number of ethical issues surrounding data privacy, security, policies and workforces. Generative AI technology can also potentially produce a series of new business risks like misinformation, plagiarism, copyright infringements and harmful content. Lack of transparency and the potential for worker displacement are additional issues that enterprises may need to address.

There is a consensus on the AI ethics guidelines that seem to converge around the principles of transparency, justice and fairness, non-maleficence, responsibility, and privacy. Despite some signs of convergence, organisations face often ill-defined and potentially contradictory guidelines, hence the need for organisational AI governance to operationalise into practice[19].

With this in mind, an organisational value alignment requires the organisation's management to state the value base and AI ethics principles on the basis of dialogue between the relevant stakeholders, and ensure adherence to these across the organisation's AI systems.

Note: (19) Artificial Intelligence: the global landscape of ethics guidelines, https://arxiv.org/pdf/1906.11668.pdf

Data & AI Risk Management

The use of Generative AI in the workplace carries several risks that organisations need to be aware of. Organisations should review and uplift where necessary their existing risk management framework to include data and AI risk with a particular focus on Generative AI.

Additional algorithm risk assessment should cover, to the extent possible, a wide range of Generative AI algorithm related risk sources and causes. As identifying biases and unfairness is often complex and contentious, the reviews should involve ethical and legal experts, particularly if the organisation intends to use the algorithm in a high-risk use case.

Internal and external risks for the successful deployment of Generative AI algorithm applications should be identified on time, mitigated, and monitored in both the development phase and the control phase.

To arrive at an acceptable Generative AI system impact, organisations should 1) conduct a thorough analysis of potential impacts the system may have on its users, subjects or affected parties, or the environment, 2) develops and implements a risk minimisation plan. The risk minimisation plan should be designed to guarantee that the AI systems are acceptable and aligned with the organisation's values and risk tolerance.

By doing so, businesses can ensure that the benefits of Generative AI are realised without compromising the integrity of their operations or their reputation in the marketplace.

Governance Approach for Organisations

Regulatory Compliance

AI systems are typically subject to a variety of regulatory instruments that may force particular design choices, constrain functionalities, or in extreme cases make implementing a specific design, use case, or business model impossible. Once a tentative understanding of the future intended use case and users of the AI system is reached, organisations should ensure that the legal function conducts an in-depth analysis of the system and its regulatory environment.

The analysis should assess regulatory risks associated with known design options, identify key design constraints, identify design areas with significant regulatory implications and, outline possible design options and their implications. These regulatory focal points should be clearly communicated to all parties active in developing or implementing an AI system within the organisation.

The AI Act[20] was proposed by the EU in April 2021 and the legislation is currently under discussion in the European Parliament. It aims to strengthen the rules on data privacy, transparency, human oversight, and accountability. It also addresses the ethical questions and implementation challenges across sectors.

Recently UK government[21] published a policy paper of a pro-innovation approach to AI regulation underpinned by 5 principles. It is imperative for organisations to track the legislation process, monitor and assess the current systems, and proactively address any compliance issues.

Data & Data-Driven Practices

Generative AI systems consume tremendous volumes of data that could be inadequately governed, of questionable origin, used without consent or contain bias. Additional levels of inaccuracy can be amplified by social influencers or the AI systems themselves.

Data is crucial to both AI systems and algorithm development and operations. The organisation should ensure that the proper documentation is in place, describing the AI system data sources, the (meta)data regarding models, input, and output; the data quality metrics, the monitoring and health check designs; the policies for storing and destruction in accordance with external and internal requirements.

From an operational AI governance, organisations must ensure that data is sourced, quality and health checked (i.e., data resources may be subject to deterioration over time), used, monitored and destroyed in alignment with both the strategic goals and values, and in accordance with external and internal requirements.

As training data, validation data, and operational data often differ qualitatively, the algorithm owner should ensure that the organisation understands the differences, designs and implements appropriate workflows and interfaces for pre-processing each data category. These practices could also safeguard against adversarial attacks, i.e., type of cyber attacks that aim to deceive AI models by inputting malicious training data that is carefully crafted to mislead.

Note: (20) Artificial Intelligence (AI) Act proposal, https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1&format=PDF
         (21) AI regulation: a pro-innovation approach, https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper

Governance Approach for Organisations

Technology of AI System

Organisations should select the appropriate tech stack and solution according to their use cases of Generative AI, ensuring that the AI system is designed, developed, operated, and monitored in alignment with the organisation's strategic goals, values and risk tolerances.

The designers and developers of the AI systems play a key role in the operational governance. They may either train and host a private model in closed on-prem environment, use a fine-tuned version in public cloud, or simply use the original version trained by the model creator. Either way, it is essential to have mechanisms in place to monitor the AI systems (which are IT systems) through their life cycle, from conception to operation and monitoring[22].

Operational AI governance components and tasks of an AI system include, but are not limited to: repository, operating environment, architecture, version control, deployment and operational metrics, performance monitoring, health check, verification and validation, approval.

Algorithms

While some algorithms are developed in-house and others procured from vendors, the organisation's understanding of the intended uses of an algorithm together with its possible misuses is key to sustainable AI development and use.

Organisations should work to develop methods for interpreting the model outputs which can include developing explainable AI techniques, enforce human review for outputs with low confidence scores, or providing training on how to interpret the outputs. Last but not least, the use of model outputs should always follow ethical guidelines to ensure that the model outputs are used in a way that is ethical and responsible.

Monitoring algorithm performance and conducting regular comprehensive algorithm health checks, must be systematic and metrics-based to achieve consistency over time, and ensure that the organisation sustains the desired level of operational performance.

Note: (22) AI governance in the system development life cycle: insights on responsible machine learning engineering, https://dl.acm.org/doi/10.1145/3522664.3528598

Conclusion

We asked ChatGPT to write a conclusion for this article using the following prompt:

"Write a conclusion paragraph for an article talking about legal and ethical challenges in using Generative AI with proposed adoption and governance approach for organisations."

Here is the response:

"In conclusion, the widespread adoption of Generative AI presents both legal and ethical challenges for organisations. The use of this powerful technology raises concerns regarding intellectual property rights, privacy, and potential biases. To navigate these challenges, organisations must establish a comprehensive governance framework that incorporates legal compliance, ethical considerations, and transparency. This framework should involve collaboration between experts from various disciplines, including AI researchers, legal professionals, and ethicists. By adopting such an approach, organisations can ensure responsible and accountable use of Generative AI, while minimising risks and maximising the benefits for both their operations and the wider society. Ultimately, it is crucial to strike a balance between innovation and safeguarding fundamental values, paving the way for a future where Generative AI can be harnessed for the betterment of humanity."

Key Takeaways:

✓ Generative AI is a disruptive technique with transformative potential across various domains.

✓ The use of Generative AI raises significant legal and ethical concerns that require careful considerations.

✓ Organisations must establish a comprehensive governance framework that incorporates legal compliance, ethical considerations, and transparency to harness its potential responsibly.
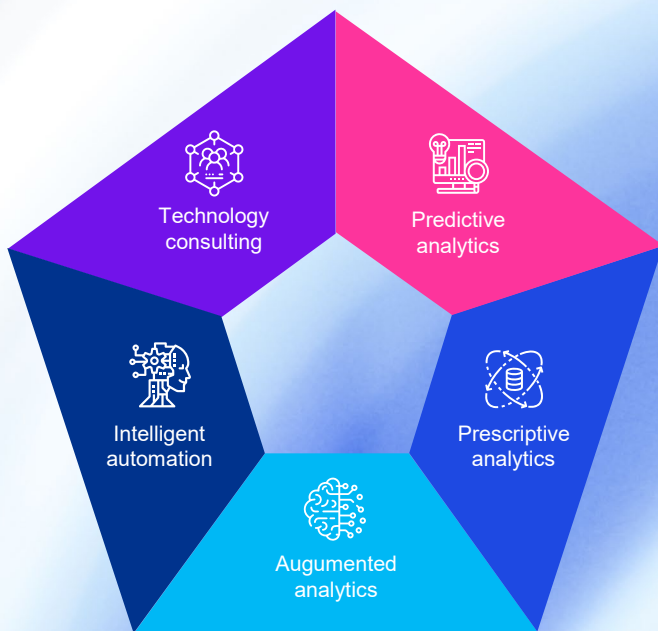
How KPMG can help you

KPMG in Poland's portfolio includes more than 100 different applications of AI in business. Our support ranges from building awareness to advising on finding and evaluating areas highly amenable to AI, implementation, scaling and maintenance. We assist in the selection of optimal technologies, platforms and standards for building AI solutions. These models are able to generate outputs that resemble human-created content, such as text or other media formats (image queries in the latest GPT-4 model) based on a user prompt. Many people believe that this technological progress marks the beginning of a new era for Artificial Intelligence and has the potential to lead the way for the fifth industrial revolution. Since its development, a wide range of applications based on Generative AI been created such as assistants for search engines, tools to enhance office productivity, helpers for coding, virtual tutors for education, and many others.

- Technology consulting
- Predictive analytics
- Intelligent automation
- Prescriptive analytics
- Augumented analytics

Benefits for clients

By integrating artificial intelligence into our operations, we can:

- ✓ generate significant savings (intelligent automation),
- ✓ improve planning processes (predictive analytics),
- ✓ increase revenue and profitability (prescriptive analytics),
- ✓ strengthen management information for better strategic management (Augmented Analytics).

Technology consulting

Leveraging scalable platforms, transparent methodologies and generating synergies with existing infrastructure allows for accelerated deployments and successful adoption of artificial intelligence within an organisation.

Project examples:

- Selection and implementation of a platform for Data Scientists.
- Preparation of implementation standards.
- Monitoring management and maintenance of Machine Learning models.
- Audit of existing solutions.

Predictive analytics

The use of state-of-the-art modelling techniques and previously unavailable data sources allows the generation of accurate forecasts that support sales, risk assessment, planning or customer location processes, among others.

Project examples:

- Preparation of a demand, sales or cost forecasting tool and integration into planning processes.
- Implementation of advanced risk models.
- Automatic prediction of customer departure or basket abandonment.
- Estimation of customer potential.

Prescriptive analytics

Price management, promotion or customer communication are a few examples of areas gaining huge opportunities by using artificial intelligence to generate answers: what to do to maximise your goal.

Project examples:

- Implementation of price recommendations.

- Construction of Marketing Mix Models (MMM).

- Automatic analysis of promotions.

- Optimisation of direct marketing.

Augumented analytics

The automatic generation of insights is one of the many benefits of using machine learning. Algorithms can automatically search for correlations, detect anomalies, signal impending shocks and assess the relevance of factors that determine a company's KPIs.

Project examples:

- Automatic detection of anomalies.

- Generation of inferences to accelerate the work of analysts.

- Intelligent Data Quality.

- Implementation of conversational data analysis mechanisms - e.g. automatic editing of summaries.

Intelligent automation

Deep learning allows a human level of efficiency in image verification and text reading. As a result, processes such as data extraction from documents, request classification or message analysis are shortened to a minimum.

Project examples:

- Implementation of automated evaluation of customer requests.

- Digitisation of documents.

- Optical quality verification.

- Branch traffic analysis and customer segmentation.

Our services are distinguished by the strong connection between business and technology. We guarantee concrete numbers that assess the success of using AI, i.e. savings and profit growth measured by a range of metrics such as:

Application handling time	Number of abandoned purchase processes	Number of stock-outs	Customer retention rate	Amount of fraud

KPMG

Contact

Leanne Allen, PhD

Partner
Data, Data Science &AI
leanne.allen@kpmg.co.uk

Tao Guo, PhD

Director
Data Science & AI
tao.guo@kpmg.co.uk

Diana Dobre, PhD

Data Science Manager
Data Science & AI
diana.dobre@kpmg.co.uk

Neha Gupta

Senior Data Scientist
Data Science & AI
neha.gupta7@kpmg.co.uk

Łukasz Dylewski

Director
Data Science & AI
ldylewski@kpmg.pl

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

kpmg.com/socialmedia