White Label Consultancy

White Paper

NAVIGATING THE RISKS OF LARGE LANGUAGE MODELS

JUNE 2023

Table of contents

1. Introduction

Artificial intelligence has advanced significantly in recent years, particularly in the field of Large Language Models (LLMs) that simulate human language and generate text. The emergence of advanced LLMs, like ChatGPT, has ignited interest among businesses, encouraging them to explore the potential benefits of implementing these innovative technologies. However, businesses must assess the legal landscape and the data protection framework that applies, especially in the European Union (EU). This whitepaper aims to assist businesses by guiding how to effectively utilise LLM technologies while complying with EU privacy and data protection requirements.

While LLM technologies have recently gained attention, the concept dates to the 1960s when the first-ever chatbot, Eliza, was created. However, this primitive chatbot relied on pre-set functions and could not generate complex outputs. LLMs have now evolved to the point where they can predict, generate, and understand human language. Consequently, they hold potential for businesses seeking to enhance efficiency. For instance, LLMs can automate responses to inquiries to provide timely information to customers. Moreover, by reducing the need for human involvement and automating manual procedures, LLMs can significantly decrease cost and increase speed.

However, the versatility and functionality of LLMs often come at a cost, including increased data protection risks due to their data-intensive nature. In the EU, privacy and data protection are principally governed by the General Data Protection Regulation (GDPR), which covers the collection, storage, and automated analysis of personal data. Therefore, businesses considering deploying LLMs, especially those relying on personal data, must prioritise compliance with the GDPR.

Failure to comply with the GDPR requirements can severely affect companies utilising LLMs. An example was when, in March 2023, the Italian Data Protection Authority (DPA) ordered OpenAI to cease data processing and ensure compliance with data protection requirements regarding its ChatGPT chatbot. Similar actions by DPAs are likely to occur in the future concerning the use of LLMs.

To lawfully incorporate LLMs into a business context, companies should adhere to data protection requirements from the outset and adopt a risk management program to mitigate risks that may arise from using these tools. This whitepaper provides an overview of the specific GDPR obligations and the critical privacy and security risks that business users of LLMs should consider when employing generative AI tools.

It is important to note that data protection obligations are not the sole regulatory requirements for using LLMs. Businesses intending to use LLMs for commercial purposes must also consider other regulatory implications, such as intellectual property, consumer, labour, and anti-discrimination laws. Additionally, when using LLMs, businesses should consider legal implications and other risks they may entail, such as quality assurance, employee productivity, and the potential incorporation of sensitive business information into future versions of LLM training datasets.

2. What are LLMs?

Before explaining the concept of LLMs, it is helpful to define Generative AI. Generative AI is a type of artificial intelligence whose

main objective is to generate artificial content. Generative AI can involve the non-human or artificial generation of images (like Dall-E), audio (like Vall-E), code (like Github Copilot), and text (like ChatGPT, LLaMA or PaLM). Following this classification, Large Language Models are a type of generative AI whose objective is to model language, producing a simplified digital representation of the language, and it is used to generate text.[1] LLMs enable the processing of tasks where text is the main input, and the output of the LLMs can be used in many different areas to automate text production or processing.

For instance, LLMs can be used in the business environment for different objectives such as:

- **Text production:** LLMs can be used to produce first drafts of any working document (e.g. emails, guidelines, procedures, newsletters), which can also be produced in a specific writing style (e.g. academic, business, creative, childlike).
- **Word processing and editing**: LLMs are capable of performing word processing tasks, which include text classification (assigning predefined categories to open-ended text), sentiment analysis (determination of the emotional tone of the message), text summary, translation, and FAQ systems (creating FAQ answers for self-service customer support). In addition, LLMs are useful to paraphrase original texts or to check grammar or spelling mistakes (like a word processing application).
- **Programming code:** LLMs can also assist software developers or architects in developing programming code, developing code from natural language and translating code into different programming code.

The use of, and ability to leverage, LLMs will inevitably differ from business to business. In some cases, LLMs may promise significant productivity or cost-cutting measures. For instance, LLM-driven chatbots may reduce customer service spend and increase customer satisfaction. In other cases, the use of LLMs may be more ad-hoc, like an employee in the marketing department using ChatGPT to help create a new company slogan or a presentation in PowerPoint about a general topic. These examples are drastically different in scale, but neither are without data protection-related risks.

3. Business using LLMs

Like any other product used within businesses that handle personal data, LLMs should undergo a thorough evaluation before implementation by businesses. While general recommendations and practices applicable to adopting new technologies are likely relevant to LLMs, certain uncertainties surround their processing operations, output generation, the potential emergence of vulnerabilities, and the persuasive nature of human-like LLM outputs. Consequently, decision-makers within organisations must thoroughly assess the opportunities and risks associated with utilising these tools – not least of all in a data protection context.

The following sections provide the main considerations for organisations using or planning to use LLMs for business purposes.

[1] See https://cset.georgetown.edu/wp-content/uploads/What-Are-Generative-AI-Large-Language-Models-and-Foundation-Models.pdf. In this whitepaper, the terms generative AI and LLM may be used interchangeably.

3.1 Understand the organisation's role and obligations

Knowing an organisation's role in the AI lifecycle and the data processing chain is crucial to determine its duties and obligations.

Regarding the AI lifecycle, the AI provider and the AI user are the two most important roles. An AI provider is the organisation that develops the AI system. In this case, the AI provider is the LLM developer, which offers the LLM to the organisation that uses it for its own business purposes. The AI user, on the other hand, is the organisation that uses or "deploys" the AI system under its authority and for its own purposes. For instance, in the case of LLMs, OpenAI, Google, and Microsoft are typical LLM providers and organisations that procure the AI system will be the LLM user. These definitions, which mirror the terminology employed by the AI Act draft, do not match precisely with the roles and definitions contained in the GDPR.

For the assignment of roles, the GDPR considers the party that decides the purposes and means for processing personal data. An organisation is a controller where it decides the means and purposes for processing personal data. This means that the controller determines the how and why of the processing operations. On the other hand, processors are organisations that process personal data on behalf of the controller. The controller instructs them how to process personal data by a contract or data processing agreement. By their very nature, processors cannot process personal data contradicting or going beyond the explicit instructions provided by the controller. This determination is important because most of the GDPR obligations are directed to controllers of personal data.

It is essential, then, to understand how these roles relate among themselves. The AI lifecycle can generally be divided into two stages: development and deployment or use of the AI system. While it is not the objective of this work to provide guidelines to manage the AI risks during the development of AI systems, suffice it to say here that during the development stage, i.e., the creation of the LLMs, LLM developers will assume the role of controllers. This is because AI providers to build the LLMs determine the means ("how") and purposes ("why") of the processing of personal data. They are fully responsible for demonstrating compliance with the data protection obligations. They may engage third-party service providers, which would be processors for these purposes. However, organisations using LLMs will generally play no role in the development stage.

During the deployment phase (the use of the LLM), organisations using LLMs will be considered controllers because they decide why and how to process personal data. For this purpose, they will engage an LLM provider who, as a service provider, will process personal data on behalf of the organisation using the LLM, thus becoming a processor.

A summary of the roles is provided in the table below.

LLM lifecycle	LLM Provider	LLM User
Development	Controller	No Role
Deployment	Processor	Controller

However, this oversimplifies how the relationships work in real-world scenarios and more complex relationships among organisations may emerge. For instance, a company using an LLM can be a processor in a B2B context (e.g. a company providing a communications platform that uses a chatbot and is powered by generative AI), and the organisation procuring the services of the first company will be considered as a controller and the

LLM provider a sub-processor. Also, an LLM provider will be considered a controller in the deployment stage if it processes personal data (provided by the LLM user) for its own purposes, such as service improvement, training of the LLM, or service performance and security.

Despite the abovementioned limitations, this distinction will help consider the obligations of businesses when using AI systems.

ORGANISATIONS USING LLMs SHOULD

1. Evaluate their role in the personal data processing activities.
2. Evaluate the terms and conditions of the processing agreements with the LLM provider.

3.2 Develop a solid foundation for the use of LLMs

Policies and procedures constitute the foundation of any privacy program and represent and define the organisation's administrative best practices. From the outset, organisations that use generative AI tools must adequately define, document, and communicate the policies and procedures regarding the use of LLMs. These policies and procedures must assign clear responsibilities to the individuals or roles within the organisation.

It is also critical to involve the management in the discussion around the permitted and forbidden uses of the generative AI tools and in approving these documents. Without management approval, individuals may think the organisation's rules on using LLMs do not apply to them.

Management should be informed not only about the benefits of using generative AI tools but also about the risks. Risks do not only include harm to individuals but also to the organisation itself, such as reputational damage, loss of customers, and information leaks. Hence, management should be provided with a clear overview of the identified risks, the evaluation of such risks in terms of impact and likelihood, and risk control measures should be proposed to mitigate these risks. Only with this information management can take an informed and documented decision and accept the residual risks of using LLMs for business purposes.

Where management approves the business use of generative AI tools, proper governance artefacts should be drafted, approved, and communicated to the organisation. For instance, a generative AI policy could be established, setting the "why" and the "what" of generative AI, establishing the organisations' goals and priorities for generative AI implementation, and providing guidance on acceptable and forbidden uses of LLMs. These policies should also consider, where relevant, elements from industry standards such as ISO/IEC 23894:2023, ISO/IEC 38507:2022 or NIST AI RMF, and consider practical examples produced by LLM providers.[2]

To complement the governance architecture created by policies, procedures should also be developed. Procedures provide the "how to" for the policies and guide the implementation of policies, specifying detailed instructions to ensure compliance with the generative AI policy. Assuming that the use of generative AI tools in a business context will entail the processing of personal data, the procedures should detail the steps the organisation will take to guarantee that the information is accurate and ready to use or disclose. Procedures can also explain how the

[2] See for instance, Microsoft's Code of Conduct https://learn.microsoft.com/en-us/legal/cognitive-services/openai/code-of-conduct, Google Policies for Generative AI https://policies.google.com/terms/generative-ai/use-policy , Wired Generative AI Policy https://www.wired.com/about/generative-ai-policy/

organisation will respond to requests from data subjects (for erasure, correction, or explanation).

Finally, the governance package should include instructions and guidance for employees to support operational tasks and foster the responsible use of generative AI tools. Developing guidance is important because it can provide employees with practical examples on how to prompt the tool, dos and don'ts,[3] suggestions on potential use cases and resources for a deeper dive into the use of the LLMs.

ORGANISATIONS USING LLMs SHOULD

1 Evaluate risks of using LLMs.
2 Inform management about the risks and control measures.
3 Obtain management approval for the business use of generative AI tools.
4 Develop a governance structure, including relevant generative AI policies, procedures, instructions and guidelines.
5 Develop a generative AI risk management framework.
6 Implement a Risk Management Policy, which includes considerations of industry standards (such as ISO/IEC 23894:2023, ISO/IEC 38507:2022 or NIST AI RMF).
7 Establish clear roles and responsibilities regarding the mapping, measuring, and managing AI risks.
8 Put in place procedures to determine the level of risk management activities based on the organisation's risk tolerance.
9 Create guidelines to identify the impacts of the LLMs, to establish risk reduction controls and to inform employees how to use the LLMs.
10 Provide guidance and best practices to teach individuals interacting with the LLM how to prompt the LLM.
11 Review the risk management procedures and its outcomes.

3.3 Build a culture of responsible AI use

Corporate culture helps organisations to navigate the regulatory complexities of today's businesses. Senior leadership should be committed to developing and embedding an organisational culture aligned not only with the organisation's strategy, mission, values and objectives but also with the regulatory environment and shared ethical values of the industry.

Organisations using generative AI tools should create an environment where values such as the protection of privacy, responsible use of AI systems and security of personal information is at the core.

In this regard, senior leadership must highlight the crucial role of ethical considerations, data protection and security when processing personal data and, in a broader sense, compliance with the statutory obligations, and organisational policies and procedures produced by the organisation.

Building a solid culture within the organisation requires, first, identifying and understanding the current state of awareness within the company related to the potential issues that may arise when using LLMs. This can be done using simple questionnaires tailored to the potential use cases of the organisation.

The results of this rapid evaluation will provide an overview of the areas and topics that need improvement. While organisations should prioritise

[3] For instance, the City of Boston has published extensive guidelines for the use of generative AI: https://www.boston.gov/sites/default/files/file/2023/05/Guidelines-for-Using-Generative-AI-2023.pdf

areas and topics where the information processed presents higher potential risks, the increase of awareness should be conducted company-wide.

Later, develop governance documents (if not yet developed) regarding the use of generative AI tools, and communicate them throughout the organisation, ensuring that all relevant stakeholders are aware of their roles and responsibilities in supporting a responsible AI ecosystem.

Another measure to increase the awareness level within the organisation is to conduct awareness-raising sessions about the use of LLMs and encourage discussion about the ethical implications of relying on these tools.

Finally, organisations should bear in mind that these are emerging technologies and the potential use cases and risks can rapidly vary. Training and awareness sessions cannot be a one-off exercise. For this reason, it is essential to track and monitor progress to improve employee performance within the organisation continuously.

ORGANISATIONS USING LLMs SHOULD

1. Identify AI awareness level within the organisation.
2. Identify areas and topics that need improvement.
3. Communicate policies, procedures and guidance related the use of LLMs.
4. Ensure employees understand their responsibilities regarding the use of AI.
5. Conduct training sessions about the use of generative AI.
6. Track and monitor the progress and effectiveness of the training sessions conducted.

3.4 Protect the confidentiality of the input data

As previously mentioned, LLMs are powerful tools to automate text-based tasks for text generation, text processing and coding. This means that, by its nature, the organisation using the LLM must share text-formatted information with the LLM provider.

At the same time, confidentiality is a core pillar of the protection of personal data, and it requires the protection of the data from unauthorised access, processing and disclosure.

Organisations should remember that data breaches do occur. In March 2023, OpenAI made a public statement informing that a bug in an open-source library allowed some ChatGPT users to see some information of other users. The information included not only payment-related information about ChatGPT Plus subscribers but also information about the content of the conversations. Regarding the content of the conversations, OpenAI informed that titles from another active user's chat history and the first message of a newly-created conversation (if both users were active around the same time) were exposed in the incident. [4] It was also reported that Samsung banned the use of generative AI tools after suggestions that an employee accidentally input sensitive internal source code to ChatGPT. [5] In this context, organisations should consider the risks associated with the input data seen as confidential (and potentially personal) data.

When an organisation uses a cloud-based LLM offered by an LLM provider (for instance, through an API, integrating the LLM to the company's applications or businesses), the deployer must send the information to the LLM provider. How the

[4] See https://openai.com/blog/march-20-chatgpt-outage

[5] See https://www.bloomberg.com/news/articles/2023-05-02/samsung-bans-chatgpt-and-other-generative-ai-use-by-staff-after-leak

LLM provider uses this input data will depend on the particular service and system. However, at minimum, the information entered in the prompt will be shared with the LLM provider to deliver the output. For instance, if someone requests the LLM to translate a document, the information contained in the document will necessarily be disclosed to the LLM provider to deliver the translation. Where the document contains personal data, data protection laws apply. This means that confidentiality is not ensured since the provision of the service entails sharing the information with the LLM provider.

In addition to this minimum level of information sharing, which is necessary to deliver the service requested by the organisation, the LLM provider may use the information collected for further secondary purposes. Privacy policies of the most popular LLM applications indicate that the LLM providers may also use the information entered into the LLM (be it personal or non-personal data) to further train the LLM and to improve the performance of the system. This also includes, for example in translation services, the corrections made by individuals using the service.

This is information that businesses should consider in advance. If the LLM provider, despite contractual restrictions, uses the data for further purposes it becomes a controller for this processing and it is responsible both before the individuals and the LLM user for the unlawful use of that personal data. LLM users, as controllers, are responsible for engaging LLM providers, as processors, that provide sufficient guarantees to implement appropriate measures to process personal data lawfully. In this case, the responsibility of the LLM user concerns vendor vetting and monitoring. For this reason, engaging a reputable LLM provider is key to avoiding compliance issues.

However, if the LLM user agrees to the further use of the personal data by the LLM provider (e.g. training the AI system), the latter will still be the controller for this purpose, but the LLM user is the main responsible for informing individuals about these circumstances and obtaining specific consent for this further processing of personal data by the LLM provider. Here, the responsibility concerns not only vendor management but also the provision of adequate information about data usage and consent collection from individuals.

Some providers of LLMs allow organisations to exercise more control over their data and offer paid versions (branded as "Business", "API", etc) which restrict any further use of the data inserted in the prompt, for instance, for service improvement or training the LLMs. In contrast, the information entered by individuals using free services is often, by default, used to improve model performance and training of LLMs.

For these reasons, it is recommended not to input personal information or commercially sensitive data (like information protected by IP, trade secrets, etc). Also, organisations should check whether the generative AI tool provides an option to turn off the prompt history to avoid input data being used for training the LLM.[6]

Additional risks can be encountered through third-party plug-ins that may be offered by LLM developers. These involve linking LLMs to third-party APIs via plug-ins to enhance the LLM's capabilities or improve its functionalities. For instance, the LLM can act as an intelligent API caller: an individual requests the LLM to make hotel recommendations for a city, and the LLM could call a hotel reservation plugin API, receive the API response, and produce the response for the

[6] See https://www.adelaide.edu.au/technology/secure-it/generative-ai-it-security-guidelines

user.[7] However, this creates additional third-party risks due to the sharing of information with third-party APIs.

Finally, organisations should consider the problems around international data transfers and the processing of personal data in jurisdictions that do not provide an adequate level of protection. Many LLM providers are located in the USA, which means that the processing of personal data will be conducted in a jurisdiction that does not have an adequacy decision (at the time of writing). Organisations should identify an appropriate transfer mechanism to compensate for the risks that originated from international data transfers. These situations create risks that should be evaluated accordingly in the business' Transfer Impact Assessment.

One alternative is selecting providers of LLMs located in the EU. However, this does not entirely eliminate the risks, since providers of LLMs may process certain aspects of the text-based tasks abroad. For instance, the translation of PDF documents using DeepL Translator. DeepL processes the data in the EU, but uses Adobe API to transform the information contained in the PDF into formatted text and then it converts the translated document back to a PDF. The transformation of both documents is processed by Adobe Inc., located in the USA.[8]

Even if the European Commission has communicated that the EU-US Data Privacy Framework will be signed before the summer, there is still uncertainty about its potential outcome. Organisations planning to reap the benefits of generative AI should consider these risks and take decisions to mitigate them.

A potential solution for cases where the confidentiality of the information is of utmost importance is the use of locally hosted LLMs,[9] like GPT-J-6, which is a language model to be hosted and used on-premises. The increased privacy – and compliance – benefits, however, come with a cost in performance.

ORGANISATIONS USING LLMs SHOULD

1. Review the data policy of the LLM provider, particularly in relation to the re-use of input data.
2. Establish governance policies restricting the kinds of input data.
3. Do not input personal data or other confidential information to public LLMs.
4. Do not submit queries to public LLMs that would lead to problems if they were made public.
5. Request users to turn off the prompt history logging if available in the tool.
6. As far as possible, use paid business versions of LLMs for business.
7. Consider the risks of developing plug-ins for connecting third-party APIs.
8. Consider the implementation of on-premises solutions or locally hosted LLMs (like GPT-J-6B).

3.5 Supervise the output of the LLMs

The biggest advantage of LLMs is at the same time one of its major weaknesses. LLMs produce convincingly human-like text, well-structured and without grammatical or spelling errors, and their

[7] See https://platform.openai.com/docs/plugins/introduction

[8] See https://www.deepl.com/en/blog/deepl-pdf-files-translation

[9] See for instance guidance from UK National Cyber Security Center https://www.ncsc.gov.uk/blog-post/chatgpt-and-large-language-models-whats-the-risk and the BSI (German Federal Office for Information Security) https://www.bsi.bund.de/SharedDocs/Downloads/DE/BSI/KI/Grosse_KI_Sprachmodelle.pdf

outputs are often supported by sources from which the information is taken. LLMs foster the impression among users that the information is accurate, leading to a tendency to over-rely on machine-generated outputs and, more generally, on the abilities of the generative AI tools (automation bias). This creates a vulnerability since those leveraging LLMs may over-rely on their outputs and underestimate other evidence. However, LLMs are prone to errors (usually called "hallucinations") which means that the text generated, albeit convincing, is flawed, biased, or simply incorrect. Unlike humans, which base their knowledge on real-world references, LLMs derive their knowledge from the text used during the training stage. So, while LLMs are very effective at creating convincing text, the outputs are not necessarily accurate.

The most common sources of errors in the output of LLMs (hallucinations) can be summarised as follows:

- **Incorrect information**: LLMs are trained based on large publicly available information taken from the internet (for instance, Wikipedia, Reddit, etc). This means that the datasets used to train the algorithm may be inaccurate or incomplete. If the LLM is prompted to generate an answer from which the majority of the data is either incorrect or incomplete, or the topic prompted was not trained at all, the output of the LLMs may be inaccurate. In addition, the training dataset may not contain up-to-date information (some LLMs, like search engines, are connected to the Internet, so the training dataset is constantly updated).
- **Biases**: The training dataset may be unbalanced, which may generate biases towards minorities or vulnerable populations. For instance, in translation, assigning masculine declensions for words like "strong" or "doctor" and feminine for words like "nurse" or "beautiful".
- **Creation of toxic or unlawful content**: LLMs can be prompted to produce toxic, offensive, upsetting speech that can be harmful to certain demographic groups.
- **Content subject to manipulative intent**: The output of the LLMs can also be subject to external attack by malicious parties (e.g. prompt injection, which hijacks an LLM's output).

Organisations should be aware of these shortcomings to address them in a proper and timely manner.

These questions are particularly relevant where the business integrates the LLMs into their own applications or solutions through an API (for instance, a business uses a chatbot connected to an LLM via an API to provide customer support to end-users or customers).

Also, organisations using LLMs to make decisions or assist in making decisions related to individuals, that produce legal or significantly similar effects on them,[10] must implement appropriate measures to protect individuals' rights. In particular, they must allow individuals to challenge the decision taken based on the LLM's output and implement human oversight on the decision. Human supervision, though, cannot be a mere rubberstamping of the LLM output. Instead, the person performing the human oversight should be competent, qualified, specifically trained, and

[10] For instance, in the context of HR analytics, see https://resources.workable.com/tutorial/llms-in-hr-analytics

provided with resources to conduct the supervision.

ORGANISATIONS USING LLMs SHOULD

1. Provide relevant information to the business' customers, employees, and/or end-users of the LLM about the limitations of the LLM. For example, which uses of the LLM are appropriate and inappropriate (e.g., confidential information), and that the LLM may produce incorrect or misleading text.
2. Inform users about the fact that, if not entirely clear from the circumstances, they are interacting with a generative AI tool (this is important when interacting with chatbots).
3. Implement a system to validate content or manual post-processing of machine-generated texts before they are used further, especially when if the LLM makes a decision that will have direct external impact.
4. Even for internal business use, verify the factual accuracy of the information delivered by the LLMs.

3.6 Assess the impacts of using LLM in businesses

Impact assessments are part of the risk management toolkit that companies must consider when using LLM for business purposes. Impact assessments are techniques that allow organisations to identify, evaluate and mitigate the potential effects that the use of an LLM by businesses can have on individuals or society.

In this particular case, organisations using LLMs for business purposes must ensure that the LLMs are used in a manner that risks are minimised to the lowest possible level.

Many different types of impact assessments can be relevant for the use of AI systems.[11] Currently, a large number of privacy laws worldwide require organisations to conduct Privacy Impact Assessments before engaging in high-risk processing of personal data. The use of AI systems, like LLMs, in general can involve many processing operations that are likely to result in a high risk for individuals.

For this reason, organisations must use generative AI tools only after performing a Privacy Impact Assessment to identify, evaluate and mitigate risks linked with the use of LLMs.

While the legal requirements of privacy laws worldwide may vary, in general, Privacy Impact Assessments require organisations to provide a detailed description of the proposed processing and purposes for which the company will process the personal information. For this activity, organisations should consider the type of information they will be processing, the types of individuals affected by the processing, the objectives they want to achieve, etc.

Organisations must also evaluate the necessity and the proportionality of the processing considering the purposes. This means that organisations must evaluate whether the envisaged processing actually achieves the purpose for the processing and whether there is another alternative, less privacy intrusive, to achieve the same goals. Later, organisations must evaluate the risks to the rights of the individuals derived from the processing. These rights include

[11] See for instance, EU Parliament AIA draft (Compromise Text 11/05/2023). Art. 29a of the AIA draft (Compromise Text) would require "deployers" of AI systems to conduct a Fundamental Rights Impact Assessment before using a high-risk AI system https://www.europarl.europa.eu/resources/library/media/20230516RES90302/20230516RES90302.pdf

privacy rights (the risk that sensitive information becomes public) but also include other rights like economic rights (disclosing credit card details can lead to fraud) or physical security (making available information about an individual's location or phone number may lead to stalking or threats).

Finally, organisations must consider and implement control measures to eliminate or mitigate the identified risks. These control measures can include technical measures (e.g. blocking access to the generative AI tool), organisational measures (e.g. developing policies, procedures and guidelines for the use of generative AI tools) and also contractual measures (e.g. selecting the appropriate contractual setup and data processing agreement with the LLM provider).

For this purpose, checking publicly available information from the LLM provider will be important. But it will be also crucial to request LLM providers all relevant clarifications and information, for instance, regarding any impact assessment already conducted by them.

Finally, risk assessments should not be written in stone. Impact assessments are living documents that should be updated at regular intervals or where there have been material changes in the factual basis upon which the assessment was made.

It is worth noting that many companies have restricted access to generative AI tools to their employees based on the consideration of the risks posed by these systems.[12] Any organisation should consider the risks and evaluate whether the identified risks are sufficiently mitigated or whether, due to the absence of satisfactory control measures, another alternative should be explored to achieve the same aim.

ORGANISATIONS USING LLMs SHOULD

1. Follow the Risk Management Framework and policies adopted by the organisation.
2. Conduct Privacy Impact Assessments following the requirements of the applicable law.
3. Review the impact assessment conducted at regular intervals or after substantial changes occurred.

4. Compliance and regulatory landscape

The previous sections have offered a comprehensive overview of the data protection framework related to the deployment of large language models in business. It's important to note that large language models continuously evolve, with new types of AI models being developed every few months. However, the legal framework surrounding artificial intelligence continues to lag behind industry advancement. Consequently, the regulation of this field heavily relies on the efforts of data protection authorities.

If businesses intend to implement LLMs in their operations, they should proactively comply with data protection requirements. Failure to do so may result in substantial fines that can significantly harm a company's finances and reputation. According to Article 83(5) of the GDPR, severe violations of data protection can lead to fines of up to €20 million or 4% of the company's total turnover from the previous fiscal year. These penalties highlight the potential magnitude of non-compliance. Therefore, it is crucial to establish appropriate technical and organisational

[12] See https://www.forbes.com/sites/siladityaray/2023/05/19/apple-joins-a-growing-list-of-companies-cracking-down-on-use-of-chatgpt-by-staffers-heres-why/

measures to ensure compliance with data protection regulations.

Furthermore, implementing robust data protection measures will also show that businesses are caring about their clients, customers, and employees. Customers are more likely to trust companies that prioritise data privacy and security. Considering the risks associated with large language models, it becomes even more vital to prioritize data protection. By proactively ensuring compliance with applicable regulations from the outset, companies can avoid potential enforcement actions in the future, and mitigate business risk.

5. Future outlook

Large language models, and the industry behind them, are developing rapidly. Thus, it isn't easy to foresee how these technologies will evolve. Despite considerable political attention, the current legal framework that governs LLMs is largely not keeping up with the development phase of those models. For example, the EU's AI Act that would cover many types of LLMs is currently amid a lengthy period of political debate. Even once enacted, the law is unlikely to apply in full until 24-36 months thereafter.

Nevertheless, there are indications that LLM providers are increasingly paying attention to data protection regulations. Their adherence to these rules should instil greater confidence in business users concerned with their own data protection obligations arising from LLM use.

LLMs will continue to provide businesses with a growing array of useful features that can enhance and make business activities more efficient. Growth areas for LLMs include content creation and customer service, which enable companies to reach new customers and retain existing ones. Additionally, LLMs offer unprecedented opportunities in market research by analysing vast amounts of textual data at a sophisticated level. As these technologies evolve further, they can help businesses adopt an even more personalized approach to cater to their customers' needs. As LLM capabilities expand, however, so does their regulatory burden. Whilst the allure of LLMs is undeniable, businesses should continue to be aware of their associated risks, including legal risks.

6. Conclusion

Large language models are rapidly evolving and they will continue to pose challenges to businesses regarding data protection compliance. This whitepaper has in-depth covered the interplay between large language models and data protection. As has been provided, businesses have a lot of various aspects which need to be considered. The businesses need to ensure that in case they deploy LLMs in their business activities, then due to the nature of these models, the data processing shall be limited to what is necessary, and the data processed shall be adequate and up to date. Furthermore, it is essential to provide efficient data security measures to protect the data that is processed. The most efficient way to stay compliant is for the companies to implement data protection by design and default, which implies enforcement of all the measures from the very start. All in all, LLMs keep evolving and it will be necessary for businesses to have an oversight of the use of these tools against data protection requirements.

7. How can White Label Consultancy Help?

While organisations can benefit from the use of generative AI tools in their business, it is essential to assess the legal landscape and particularly the data protection framework that may be applicable.

White Label Consultancy can help you to navigate the complex legal and regulatory environment to use generative AI tools in a manner that reduces compliance risks and assist your organisation in developing and using responsible, reliable, and secure AI solutions. We also adopt a responsible approach when evaluating ethical, corporate governance, and security aspects surrounding AI and machine learning technologies for our clients.

White Label Consultancy allows you to put your privacy compliance and organisation's reputation into the safe hands of experienced and senior privacy professionals. As a boutique data protection and privacy consultancy with highly specialised privacy professionals, White Label Consultancy has vast experience implementing privacy management programs across several large multinationals, working in both in-house and consultancy roles.

We combine excellent legal proficiency with deep technical knowledge and significant operational experience to offer a made-to-measure, proven, and functional approach to solving the issues that organisations face.

We can assist organisations with their DPO needs and fast-track the data protection journey towards the desired level of maturity and compliance.

Based on the individual needs of the organisation, White Label Consultancy can offer:

- an on-demand team of privacy professionals to manage your privacy matters,
- an independent resource with no conflict of interest,
- obtaining ongoing support for daily privacy tasks and complex projects,
- a data privacy assurance service,
- drafting and implementing key policies and procedures,
- training your employees and raising privacy awareness in your organisation,
- supporting you with data subject access requests and data breaches,
- acting as your appointed DPO,
- training and ongoing back-office support for your designated DPO,
- to manage communications with authorities, processors, and data subjects,
- to assist in the development and maintenance of your record of processing activities,
- to conduct the data privacy annual assessment,
- to update, maintain and implement compliant privacy policies and procedures,
- negotiate privacy clauses in contracts with vendors,
- monitor ongoing compliance with data protection requirements,
- to assist the organisation in performance of data protection impact assessments.

Check the complete suite of services we can provide to your organisation to navigate the risks of Large Language Models here

In case you have any questions, please reach out to us using the Contact form available here or via email at hello@whitelabelconsultancy.com.

CHECKLIST
USERS OF GENERATIVE AI TOOLS

Understand the organization's role and obligations		
1	Evaluate the role in personal data processing activities.	☐
2	Evaluate the terms and conditions of the processing agreements with the LLM provider.	☐
Develop a solid foundation for the use of LLMs		
1	Evaluate risks of using LLMs.	☐
2	Inform management about the risks and control measures.	☐
3	Obtain management approval for the business use of generative AI tools.	☐
4	Develop a governance structure, including relevant generative AI policies, procedures, instructions and guidelines.	☐
5	Develop a generative AI risk management framework.	☐
6	Implement a Risk Management Policy, which includes considerations of industry standards (such as ISO/IEC 23894:2023, ISO/IEC 38507:2022 or NIST AI RMF).	☐
7	Establish clear roles and responsibilities regarding the mapping, measuring, and managing AI risks.	☐
8	Put in place procedures to determine the level of risk management activities based on the organisation's risk tolerance.	☐
9	Create guidelines to identify the impacts of the LLMs, to establish risk reduction controls and to inform employees how to use the LLMs.	☐
10	Provide guidance and best practices to teach individuals interacting with the LLM how to prompt the LLM.	☐
11	Review the risk management procedures and its outcomes.	☐
Build a culture of responsible AI use		
1	Identify AI awareness level within the organisation.	☐
2	Identify areas and topics that need improvement.	☐
3	Communicate policies, procedures and guidance related the use of LLMs.	☐
4	Ensure employees understand their responsibilities regarding the use of AI.	☐

5	Conduct training sessions about the use of generative AI.	☐
6	Track and monitor the progress and effectiveness of the training sessions conducted.	☐
Protect the confidentiality of the input data		
1	Review the data policy of the LLM provider, particularly in relation to the re-use of input data.	☐
2	Establish governance policies restricting the kinds of input data.	☐
3	Do not input personal data or other confidential information to public LLMs.	☐
4	Do not submit queries to public LLMs that would lead to problems if they were made public.	☐
5	Request users to turn off the prompt history logging if available in the tool.	☐
6	As far as possible, use paid business versions of LLMs for business.	☐
7	Consider the risks of developing plug-ins for connecting third-party APIs.	☐
8	Consider the implementation of on-premises solutions or locally hosted LLMs (like GPT-J-6B).	☐
Supervise the output of the LLMs		
1	Provide relevant information to the business' customers, employees, and/or end-users of the LLM about the limitations of the LLM. For example, which uses of the LLM are appropriate and inappropriate (e.g., confidential or sensitive information), and that the LLM may produce incorrect or misleading text.	☐
2	Inform users about the fact that, if not entirely clear from the circumstances, they are interacting with a generative AI tool (this is important when interacting with chatbots).	☐
3	Implement a system to validate content or manual post-processing of machine-generated texts before they are used further, especially when if the LLM makes a decision that will have direct external impact.	☐
4	Even for internal business use, verify the factual accuracy of the information delivered by the LLMs.	☐
Assess the impacts of using LLM in businesses		
1	Follow the Risk Management Framework and policies adopted by the organisation.	☐
2	Conduct Privacy Impact Assessments following the requirements of the applicable law.	☐
3	Review the impact assessment conducted at regular intervals or after substantial changes occurred.	☐

Authors

Nicholai Pfeiffer
Managing Partner
White Label Consultancy
np@whitelabelconsultancy.com

Magdalena Goralczyk
Partner
White Label Consultancy
mg@whitelabelconsultancy.com

Federico Marengo
Senior Consultant
White Label Consultancy
fma@whitelabelconsultancy.com

Peter Davis
Special Counsel
White Label Consultancy
pda@whitelabelconsultancy.com

Norman Aasma
Junior Associate
White Label Consultancy
noa@whitelabelconsultancy.com

White Label
Consultancy