CONCLUSION

Careful and safe deployment of artificial intelligence. Discover how to organize this using 6 focus areas.

CONCLUSION Whitepaper







1	What is responsible AI?	page 4
2	How do you set up responsible AI?	page 7
3	Responsible AI in your organisation?	page 16

Introduction

Responsible AI harnesses the power of artificial intelligence, carefully and safely

Your organization wants to use artificial intelligence (AI) for strategic goals. However, you want to do so ethically and with an eye for all interests involved. And with awareness of questions such as 'what do we actually want with technology?' and 'what impact do we allow to have this on people and the planet', in addition to the questions 'what can we do with it?' and 'what is allowed and what not?'. **Ensuring that AI is used carefully and safely** is known as responsible AI.

Responsible AI is more and more becoming an integral part of organizations' business operations. It enables you to manage risk while leveraging the power of Al.

This is why:

- in your organization.
- models).
- organization's carbon footprint.

Partners and investors are more likely to do business with organizations that use technology with due prudence.

Responsible AI means, among other things, being transparent about how you process data and make decisions. By avoiding prejudices (bias) and ensuring explainability, you increase trust

Legislation and regulations, such as the GDPR and EU Artificial Intelligence Act, impose increasingly stringent requirements on the development and use of AI. Responsible AI helps to ensure compliance. In addition, responsible AI helps prevent incidents and reputational damage (for example due to discriminatory AI

Failure to comply with legal standards can lead to fines and lawsuits. Examples include fines for privacy violations or unlawful use of AI. By getting started with responsible AI now, you are preparing for stricter regulations in the future.

Sustainable AI helps you to reduce and manage your

Responsible AI is essential for organizations. Only then will you be able to build trust, minimize risk and stimulate innovation.

Organizations that put responsible AI at the centre of their operations are more profitable, more sustainable and better prepared for the future.

What responsible AI entails is subject to continuous change. Changes in social norms and values and in legislation and regulations are often a response to incidents and new insights.

In this whitepaper, we share our knowledge and experiences with you. You will read about our 'Responsible AI Framework', with practical recommendations in six focus areas. This framework enables you to permanently embed responsible AI within your organization.

So get started with responsible AI. Seize the opportunities and minimize the risks, now and in the future.



s verantwoorde ai?

CONCLUSION | Whitepaper Verantwoorde AI

What is RESPONSIBLE AI?

Responsible AI is about developing, using and managing AI in a way that is ethical, fair, transparent and explainable, in line with social norms and values. This goes beyond making your AI systems compliant with the law, such as the GDPR and EU Artificial Intelligence Act or the use of smart AI models.



1 What is responsible AI

Responsible AI is not a technical problem in its own right or a matter of compliance. It's an ethical and social issue. One in which companies, users, technology experts and policymakers together contribute to the careful use of AI.

The goal is to get the most benefit from AI, while minimizing risks.

Human control is crucial in this, because humans always carry ultimate responsibility for the results that an AI model generates and for managing the risks associated with AI use.

The dangers of IRresponsible AI

The impact of AI on our society, economy and personal lives is growing. Insufficient attention to responsible development, use and deployment of AI can lead to incidents in the areas of:

1.Equality

2.Safety Inadequate protection of personal data can lead to misuse or data breaches.

Training and using complex models, especially for generative AI, consumes enormous amounts of electricity and water. In addition, human labour involved in the AI supply chain is often an invisible and undervalued factor. Exploitation of people is a real risk.

AI can increase prejudice and discrimination through bias in data.

3. (Social) sustainability

The dangers of responsible AI

Setting up responsible AI requires a combination of technical, procedural and organizational measures. Together, these measures enable you to align responsible AI with new developments and and keep it aligned. Several aspects are important:

- Ethical design: responsible AI is developed based on clear ethical values, such as sustainability, fairness and safety.
- **Transparency**: the operation of responsible AI systems is understandable and explainable. As an organization, you are transparent about the way you use AI and how AI supports decision-making.
- Equal treatment and inclusion: responsible Al systems do not reinforce discrimination or exclusion. This requires an effort to make datasets and algorithms inclusive and representative.
- **Privacy protection**: responsible AI respects and protects personal data in which applicable laws and regulations such as the GDPR and the Artificial Intelligence Act serve as the lower benchmark.

- **Sustainability**: applying responsible use of Al does not unnecessarily burden the environment, thanks to sustainable development practices and energy-efficient infrastructure and use.
- **Safety**: responsible AI systems are robust, reliable, and designed to minimize errors and misuse.
- Human control: people remain able to influence important decisions. Al is not a replacement for human responsibility.
- **Accountability**: organizations can account for the way they use AI and the impact thereof. There are protocols in place if something goes wrong.

How do you set up RESPONSIBLE AI?

The possibilities of AI are growing each day, with systems operating more autonomously and smarter. Yet at the same time, we want to develop and use AI systems ethically and responsibly. Conclusion's 'Responsible AI Framework' offers your organization guidance and recommendations for this.

CONCLUSION | Whitepaper Responsible AI

The framework contains six focus areas. Together, these form the practical 'building blocks' for implementing responsible use of AI. The framework is a blueprint. Setting up responsible AI is always a customized process.

Responsible AI framework

Quality Assurance & human involvement

Transparency & accountability

Setting up responsible AI is an ongoing process. It involves combining technical innovation and compliance with ethical reflection.

An adequate approach must always be proportionate and appropriate: in proportion to the use of AI and its risks.

The responsible AI-framework

What recommendations do we make for responsible AI? Discover the measures in six focus areas.

FOCUS AREA 1 **Policy & governance**

The basis for responsible AI is a substantiated *policy*. This describes your strategic goals and establishes what and how AI will and will not be used. This way AI projects run consistently and in accordance with the applicable guidelines, with an eye for ethical and legal obligations. The policy must be developed along the lines of the other five focus areas.

Framework of core values

Part of your AI policy is a framework of core values that form the basis of AI development and use in your organization. Commonly used values are fairness, transparency, privacy, security and sustainability. Responsible AI is always closely linked to social norms and values. It's good to be aware that these differ by culture, era and even organization. Take privacy for example: in Europe this is highly valued. In some other regions this is different, especially when security or economic interests are at stake.

Social norms are constantly evolving. What's acceptable today may no longer be acceptable tomorrow and vice versa. In addition, different stakeholders (companies, governments, individuals) often have different ideas about responsibility. An AI policy is therefore never 'finished'. Effective policy takes into account the diversity of insights, interests and values and is regularly adjusted.

Finally, we evaluate an AI system used in a hospital differently than a system for online advertising. What is deemed responsible therefore also depends on the context in which you apply AI.

In 2024, the AI journalist 'AiFred' was taken offline the very day after its launch. The initiators discovered that AiFred gave biased responses about vaccinations, among other things, such as accusing 'big pharma' of profit-seeking at the expense of children.

2 How do you set up responsible AI?

Supervising the implementation

Policy and governance go hand in hand. Governance defines the roles, responsibilities and procedures involved in the implementation of the AI policy. How do you shape this in your organization? Monitoring policy implementation is crucial to manage risks. In addition, it provides checks and balances between the interests at play within and outside the organization.

In 2019, it was revealed that Apple Card was giving women lower credit limits than their male counterparts with similar or lower incomes and credit profiles.

CONCLUSION | Whitepaper Responsible AI

Five practical tips for responsible AI:

- Set up an AI ethics committee of ethicists, engineers, lawyers and users. They assess AI projects and monitor their ethical impact.
- 2. Designate who is responsible for compliance with the ethical guidelines.
- **3.** Establish mechanisms to monitor the impact of AI and report abuses.
- **4.** Involve diverse groups (including users, external experts and civil society organizations) in the design and implementation of AI.
- 5. Establish systems for users to report complaints or concerns about AI use (feedback mechanisms).

FOCUS AREA 2 Justice & legality

This focus area concerns compliance with ethical and legal frameworks and any other standards that are central to your organization.

Whereas legality is about what you have to do (legislation), *justice* is about what is desirable (moral action). The assessment of the latter requires a balance between costs and benefits, with an eye for the risks involved. Responsible AI is not just about preventing incidents, but also about creating trust in technology.

In 2016, Cambridge Analytica used AI analytics tools to collect personal data from Facebook users, without their consent. This data was used to create political advertisements and influence voting behaviour during the US elections and the Brexit referendum. The legal frameworks describe how your organization deals with legislation such as the GDPR and the AI Act. The ethical frameworks determine what your organization stands for and contribute to a careful weighing of interests when developing and deploying AI. Risk classification of AI models plays an important role in this.

Perform a risk analysis before implementing an AI system. Think about risks such as discrimination, privacy implications and safety risks. Always thoroughly test AI systems in different scenarios to see how they behave in practice.

Actively minimizing bias is a basic principle in responsible AI. It is therefore important that datasets are representative of all relevant groups. And that you use techniques to detect and correct bias in data. Also consider data audits: i.e. regular checks on the quality and integrity of data.

In 2024, a total of 97 African AI workers sent a letter to President Biden, demanding better treatment. They accused Meta and OpenAI, among others, of "modern slavery" due to low wages, traumatic working conditions and unexpected dismissals without compensation.

Socio-ethical aspects

While AI offers enormous benefits, the human labour it supports can sometimes be undervalued and poorly rewarded. This points to the need to also pay attention to the socio-ethical aspects of AI, with companies and policymakers joining forces to distribute the benefits of AI as fairly as possible.

FOCUS AREA 3 Quality assurance & human involvement

This focus area includes all the measures your organization takes to ensure AI models do what they are supposed to do (and refrain from what they are supposed to refrain from doing). How do you ensure the accuracy of models, improve the security of algorithms and prevent biases in data?

'Privacy by design' is a technical aspect that plays a role in **quality** assurance. This means that privacy protection is integrated into the design of AI systems. This contributes to responsible use. Minimizing data collection and anonymizing it where possible is also important, as is protecting AI systems from cyber attacks and manipulation.

Just as important is *human involvement*: control over Al output is necessary, especially when dealing with high-impact applications. This way people remain an essential link in the decision-making process.

Diverse and inclusive decision-making is crucial as well. Subjectivity is better managed by involving diverse stakeholders in the development and use of AI. Collaborations between companies, experts, policymakers and users also contribute to this.

Monitoring

Finally, continuously monitor the performance of your AI systems to ensure they continue to meet ethical and legal standards. This allows you to conduct regular audits to check the effective management of risks and use feedback to improve AI systems.

In 2023, an MIT student of Asian descent used the AI tool Playground to make her profile picture more professional. Consequently, the colour of her skin was made lighter and her eyes were coloured blue.

FOCUS AREA 4 **Transparency & accountability**

The EU Artificial Intelligence Act strikes a balance between flexibility and uniformity by providing a framework that allows for ethical flexibility, yet that sets minimum standards at the same.

Transparency is the keyword. After all, AI developers and users being transparent about their ethical choices builds trust and allows others to provide feedback. Transparency is about the information you proactively provide about the role of AI within your organization. For example, via an algorithm register or in the privacy statement.

Transparency starts with documentation: keep track of how AI models were developed and trained. And generate reports on the performance, limitations and risks of AI systems. Finally, 'Explainable AI' are systems where users can understand the logic behind decisions.

Decisions taken

Accountability goes beyond transparency. This means that the organization remains responsible for the use of AI. Management can explain why decisions have been made, taking into account all interests, as well as legislation and regulations.

As AI systems become more sophisticated and widespread, accountability becomes increasingly important. Organizations that create and/or use AI systems must be able to demonstrate that they take responsibility for the impact this has on people and society.

For example, by being able to demonstrate through audits what measures they take to limit and control risks, and how they monitor compliance with these measures.

As part of this approach, subjective standards can be supplemented with objective information. For example, by using research and data to assess which choices are helpful. Responsible AI systems must be continuously evaluated and improved.

In 2024, a German journalist checked his name in Microsoft Copilot. This tool then claimed, among other things, that he had been convicted of child abuse and exploitation. The journalist had written many articles about abuse, violence and fraud and Copilot apparently associated this with the author.

FOCUS AREA 5 **Ecological impact**

Generative AI consumes a lot of power, and the servers on which models are run use a lot of water and energy to cool themselves. Therefore, include sustainability measures from as early as the design phase, thus minimizing the *ecological impact*.

The CO_2 emissions from training large neural networks range from tens to hundreds of tons, depending on factors such as size, hardware efficiency and the energy source used by the data centre.

An article from the <u>Delft University of Technology</u> states that training the ChatGPT 3.5 model caused the emission of approximately 500 tons of CO₂. In addition, an article from the <u>Centre for Work and Income</u> stated that training a large AI model like ChatGPT produces an annual ecological footprint comparable to that of sixty people in a Western country.

A Google search uses approximately 0.0003 kWh of energy. A similar prompt in GPT-4 is about 0.0005 kWh, over 60% more. Sustainable use of AI starts with the choice of AI models: using AI models that require less computing power and data and that are modular and reusable is preferred. Also consider optimizing model and hardware architecture to limit energy consumption and purchasing software and hardware that meets environmental criteria.

Collaborating with sustainable partners also contributes to the responsible use of AI, such as data centres running on renewable energy. And implement standards for sustainable AI, such as CO_2 compensation as offset for energy-intensive AI projects.

The use of AI also offers potential for environmental gains. For example, by scheduling AI calculations at times when energy demand is low or when sustainable energy is available, and by performing those calculations closer to the user, where possible.

Finally, be critical of excessive use of AI: only use AI when it adds measurable value, and use smaller and simpler models if you can achieve the same result.

AI applications are hosted in data centres that use water to cool. In 2027, this will require approximately 4.4 to 6.6 billion m² of water. For comparison: annual water consumption in the Netherlands as a whole amounts to approximately 10 billion m².

FOCUS AREA 6 Awareness & knowledge retention

To use AI responsibly, everyone in the organization must understand what it is and what it is not, what added value it offers and what its limitations are. This calls for *awareness* and a general level of basic knowledge, plus additional training for developers, supervisors, directors and managers.

In October 2023, a team of researchers from the University of Zurich raised serious concerns about AI chatbots. It turned out that they were able to deduce very personal information such as ethnic background, location and profession from seemingly innocent chats. Organizations that develop and/or use AI solutions must pay sufficient attention to privacy and ethics. This requires a broad approach to creating awareness and skills among employees.

On the one hand, you train employees who develop and/ or implement AI in ethical considerations, bias detection and transparency practices. On the other, you educate users about how AI works, what its limitations are and how to use AI responsibly (AI literacy).

By *actively safeguarding knowledge* and continuing to develop, you create a culture in which AI can be used effectively, safely and responsibly.

WHITEPAPER AI LITERACY

Also read our whitepaper on AI literacy, an important condition for ensuring responsible AI.

Al literacy represents the knowledge and skills needed to:

- explain and understand AI;
- handle Al; and
- assess the outcomes and impact of AI.

portun

CONCLUSION | Whitepape

Our consultants contribute to responsible AI in your organization. With a wealth of experience, we realise and safeguard an optimal situation for all users and stakeholders. We do this by offering an appropriate and proportionate solution. This always involves a customized approach: the best solution depends on many factors and must fit seamlessly within the organization.

CONCLUSION | Whitepaper Responsible AI

Responsible AI IN YOUR ORGANIZATION?

Together we prepare a plan to ensure responsible AI. This way you will comply with all legal and ethical standards, now and in the future. Responsible AI is a serious topic. Therefore, start by requesting our 'responsible AI scan', so that you can discover what Conclusion can do for your organization. Our advisor will take a critical look at all aspects of responsible AI within your organization. Based on the results, we determine your ambitions and draw up a 'roadmap' to realize those ambitions, together.

Our approach in 3 steps

STEP 1 Intake with responsible AI scan

It's important that we gain an overview first. As part of the intake we will map out your specific objectives, training needs and target groups.

The responsible AI scan provides instant insight into the development needs of your organization. Based on the results, we establish concrete objectives and deliverables and draw up an action plan.

STEP 2 Responsible AI approach tailored to your needs

of the Responsible AI framework:

- Policy & governance Justice & legality Quality assurance & human involvement Transparency & accountability Ecological impact

- Awareness & knowledge retention

Based on the results of the scan and the vision and ambition of your organization, we formulate an approach to safeguard responsible AI within the organization.

The Responsible AI approach follows the components

STEP 3 Dustainable knowledge development

Our customized train-the-trainer program produces sustainable knowledge development. Initially, Conclusion provides the training, after which we train internal trainers to continue the program independently.

By actively safeguarding and continuing to develop knowledge, you create a culture in which AI can be used safely and effectively.

By actively safeguarding and continuing to develop knowledge, you create a culture in which AI can be used safely and effectively.

Want to know more or schedule an intake?

Contact Conclusion by calling 030 219 38 00 or send an e-mail to <u>info@conclusion.nl</u>. We will be happy to make an appointment for your intake.

About Conclusion

Conclusion has over a decade of experience in big data and AI and has been a member of the NEN standards committee for AI and big data for over five years. Thanks to our approach to AI literacy and competency development, we have a proven track-record of providing companies and institutions with a solid foundation for AI adoption in their respective organizations.

LEAD THE CHANGE BEFORE CHANGE LEADS YOU

Get in touch at: info@conclusion.nl

This white paper was made possible thanks to the assistance of the following experts in the Conclusion ecosystem:

Adil Bohoudi LinkedIn: <u>Adil Bohoudi</u>

Friso Spinhoven LinkedIn: <u>Friso Spinhoven</u>

