

Responsible AI

A beginner's guide

 **Direct Digital**
Holdings

 Colossus SSP®  Orange 142®



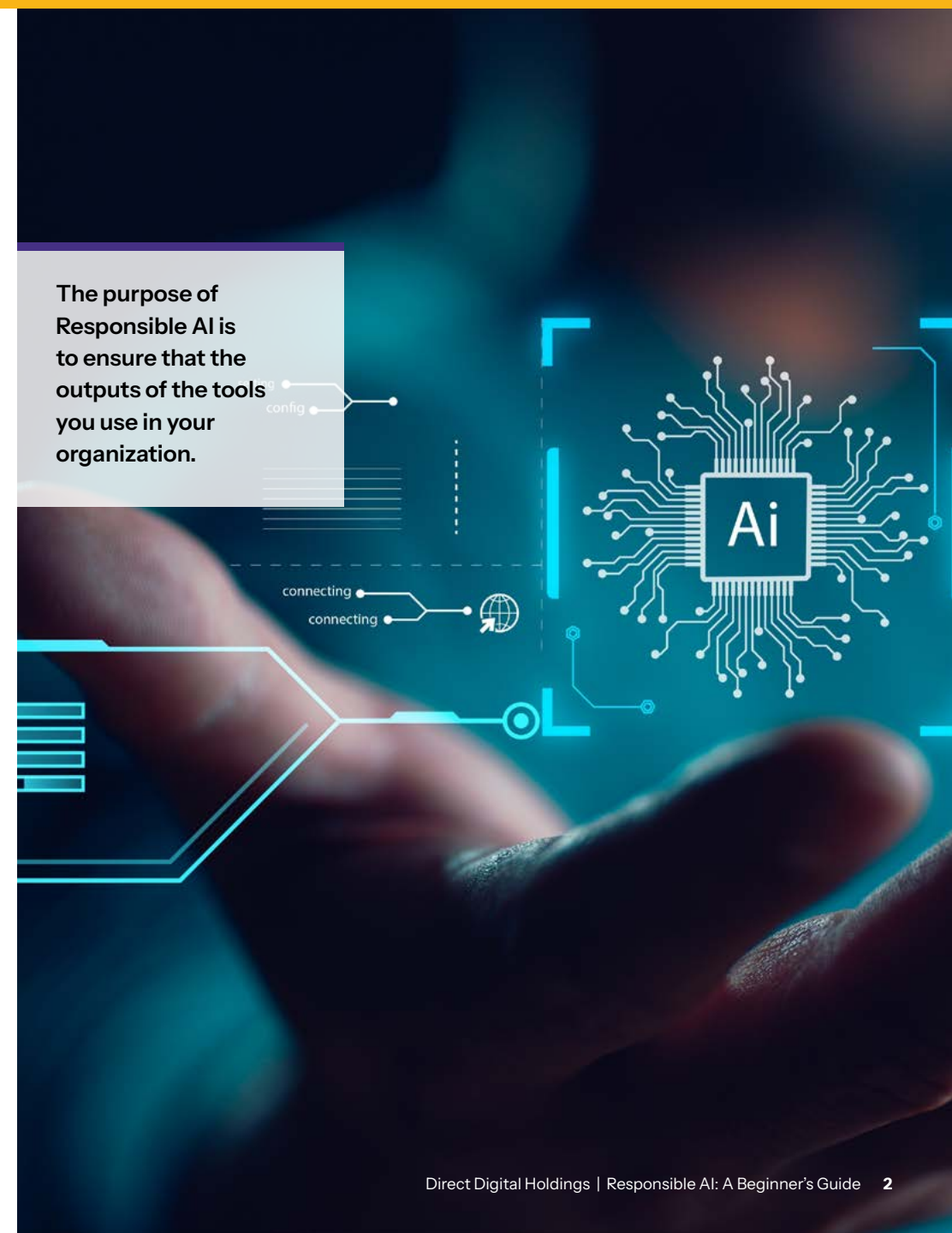
Forward

AI is changing the way we live and work, and its outputs can have big impacts on our lives. With such power comes great responsibility. If we rely on AI to decide who to call in for a job interview, or to determine the cause — and apt treatment for — someone's illness that AI needs to be trustworthy.

What started as discussions about AI ethics has grown into its own field, complete with university programs and dedicated experts in companies.

The purpose of Responsible AI is to ensure that the outputs of the tools you use in your organization are:

- Unbiased, meaning they don't treat one person worse than another (e.g. favor male candidates over female in an HR screening tool)
- Safe, meaning outputs don't cause harm to individual people or vulnerable populations or violate a users' right to safeguards that protect their data.



We can't assume that AI is completely inherently unbiased. AI is trained on data labeled (aka "annotated") by people who make decisions based on their life experiences. Its model design has been highly influenced by data engineers who, as people, have specific worldviews.

Ensuring fairness means your organization must meet very specific requirements, some of which are required by law. For instance:

- Can you explain how your AI makes decisions?
- Do users understand its decision-making processes?
- Who within your organization is accountable for the decisions?
- Is there an escalation path to report or challenge outputs?
- Who reviews your AI application on a regular basis to check for bias that may develop?

Important Note

The DDH AI Council created this guide to introduce you to the principles of Responsible AI. It does not offer a detailed, Responsible AI framework (plenty of them are readily available from industry giants and government bodies, including Microsoft, NIST, the EU, and others).

Rather, its purpose is to make you aware of the ways that AI can lead to harm and the types of actions organizations must take to protect the users and your brand reputation.

We strongly recommend you consult with legal, compliance, and data privacy experts to ensure adherence to applicable regulations and the safeguarding of organizational interests. All information is provided "as is," without warranties of any kind, and should be used at your discretion.

About the DDH AI Council

The DDH AI Council was founded to address a growing concern: the widening divide between organizations that embrace generative AI and those that are hesitant to adopt it. Generative AI is rapidly reshaping how we work, raising the overall caliber while enabling teams to innovate faster. We understand that for many business leaders, generative AI is still an unknown technology that comes with many risks. Our goal is to demystify generative AI, and to provide the education and insights business leaders need to build a roadmap for its adoption, with full confidence that its use will be safe and transformative.

Table of Contents

What Makes AI “Responsible”?	5
Why Pay Attention to Responsible AI?	7
Risk-Based Categorization of AI Systems	12
Key Pillars of Responsible AI	15
Existing Responsible AI Frameworks	18
Responsible AI Governance	20
Parting Thoughts: Integration Across Your Governance Framework	23
Glossary of Terms	25

Direct Digital Holdings is a fast growing, efficiency-focused solutions provider in the digital marketing and advertising sector. We are a family of brands serving direct advertisers, agencies, publishers, and marketers.

A woman with glasses and a dark blazer is looking at a laptop. The background is a blurred office setting with a window showing sunlight. Overlaid on the image are several semi-transparent blue boxes containing code snippets and a large, stylized brain graphic composed of many small white dots, with some parts of the brain highlighted in a lighter blue. The overall theme is artificial intelligence and technology.

What Makes AI “Responsible”?



AI errors can have serious consequences. When [Air Canada's chatbot](#) provided false information to a user, the user successfully sued the airline. As AI increasingly influences healthcare, employment, and financial decisions, proper safeguards become critical and, in many cases, a legal requirement.

Responsible AI demands key protections, including:

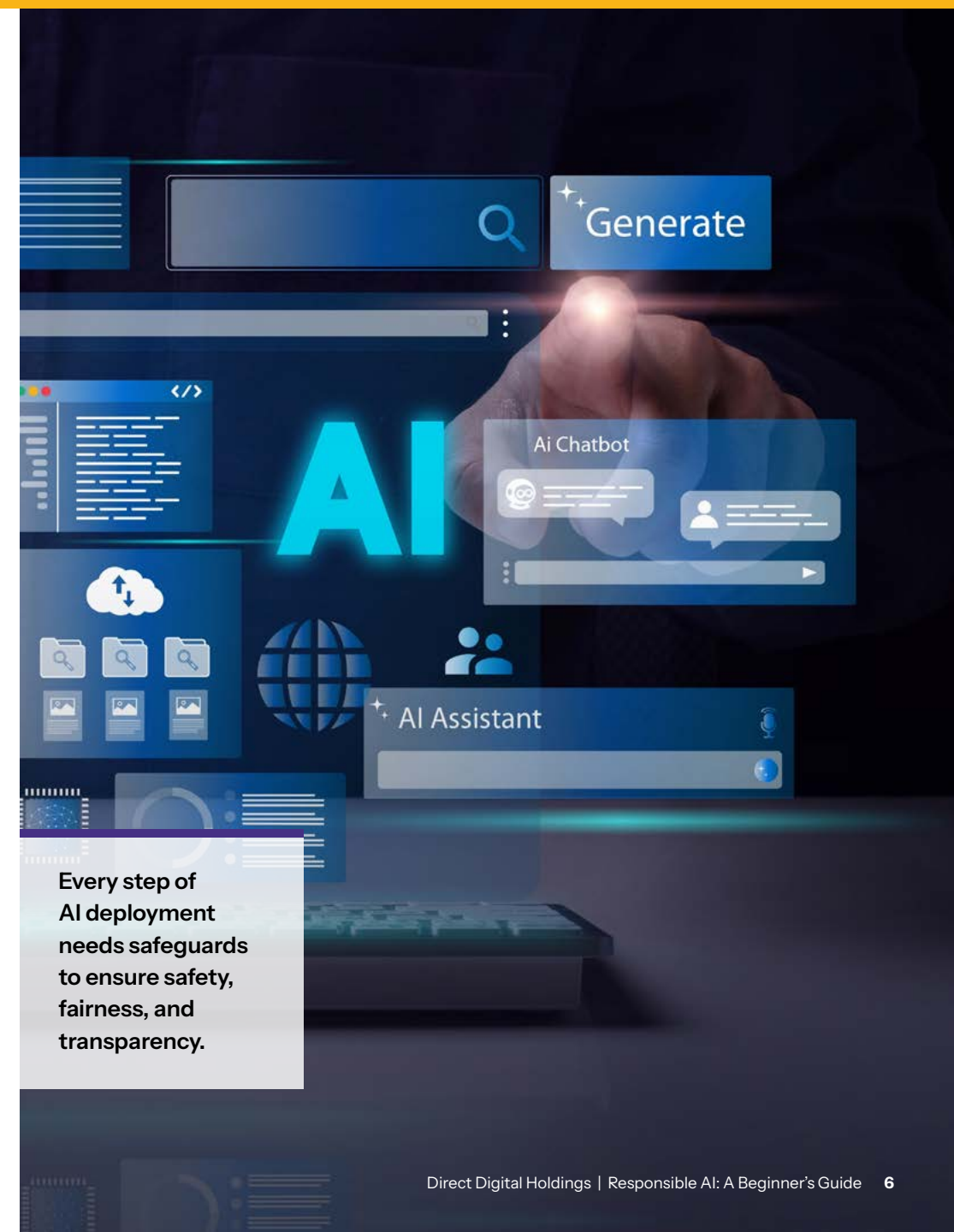
- Active monitoring and controls
- Clear accountability
- Regular bias audits
- Data privacy protection
- Transparent decision-making

Every step of AI deployment needs safeguards to ensure safety, fairness, and transparency. With millions potentially using each AI tool, bias or errors can lead to widespread harm.

Responsible AI is a framework designed to:

- Minimize risks to the both your users and your brand reputation
- Protect user privacy
- Maintain brand trust
- Benefit users, businesses, and society

Think of it as AI with guardrails, giving users confidence that the technology will help, not harm.



Why Pay Attention to Responsible AI?



All companies want to ensure their AI tools cause no harm to their users or brand reputation. As a business leader who takes pride in your work, you are naturally interested in ensuring your AI tools do not create biased outputs, violate privacy, or lead to unethical outcomes.

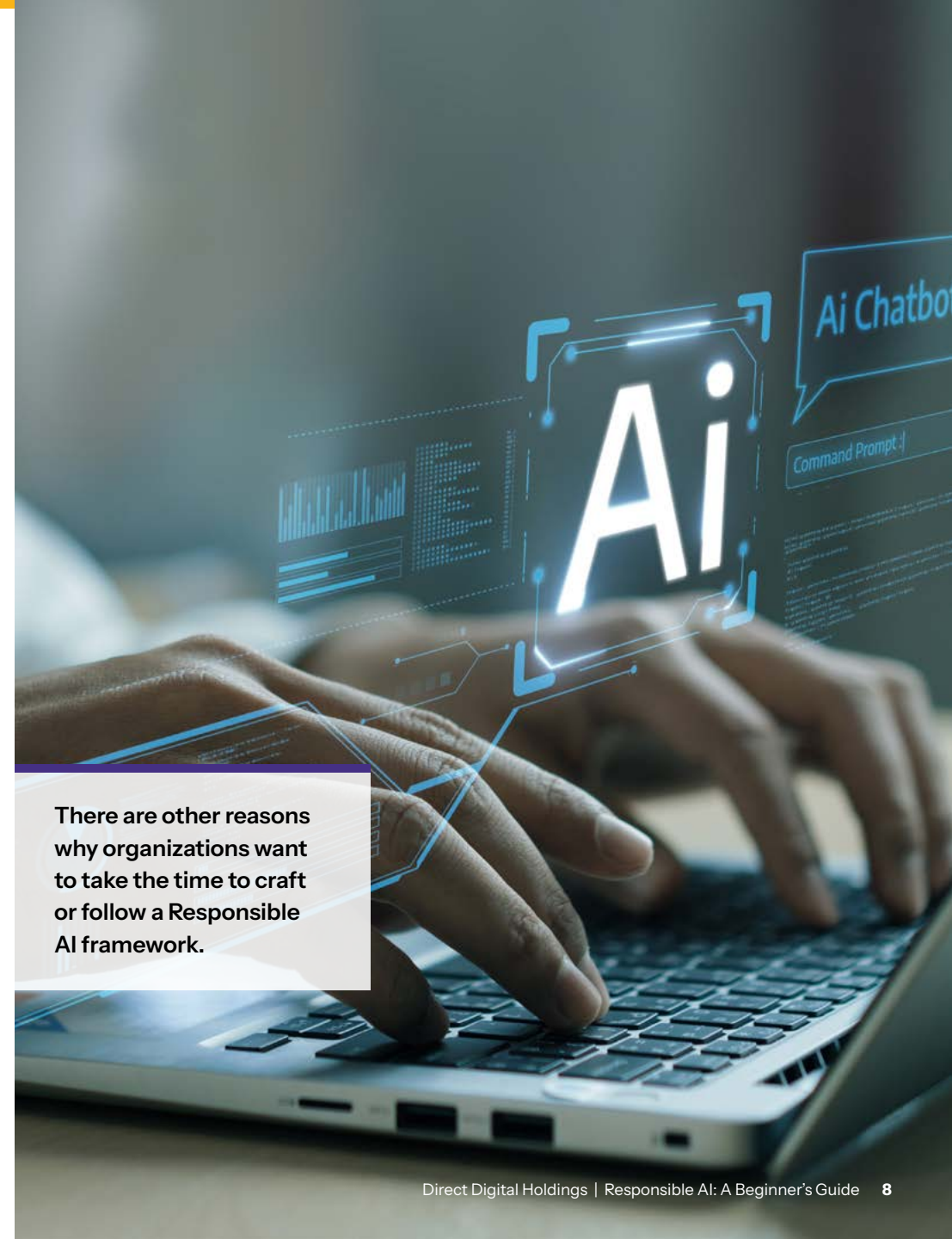
There are other reasons why organizations want to take the time to craft or follow a Responsible AI framework, which we discuss below.

Regulatory Compliance

We've mentioned earlier that in some cases you are legally required to ensure your AI is responsible. Let's look at some of the regulations that may apply to your company's AI application.

US Federal Level

As of December 2024, there is no national law governing AI, but several Executive Orders and regulatory measures either strongly advise or require it. For instance, in September 2024, the FTC launched a law enforcement sweep targeting companies using [AI for deceptive or unfair practices](#).



There are other reasons why organizations want to take the time to craft or follow a Responsible AI framework.

In October 2023, the Biden Administration issued [Executive Order 14110](#), the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. This EI sets standards for AI safety.

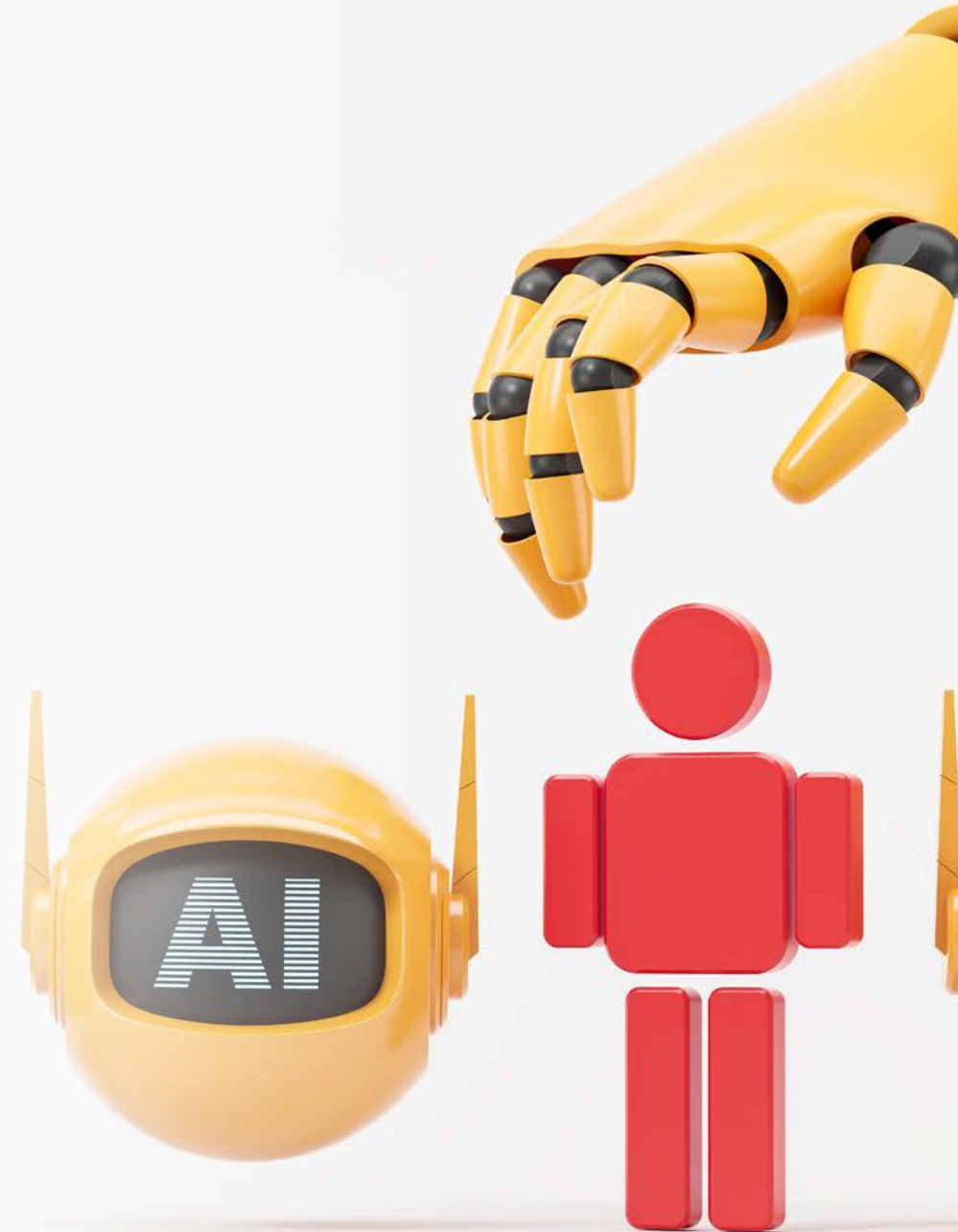
That same month, the Administration issued an Executive Order outlining a [Blueprint for an AI Bill of Rights](#), explicitly addressing algorithmic discrimination and consumer privacy.

US State Level

On the state level, legislative activity is similar to privacy regulations. Specifically, while most states have yet to adopt an AI law, all have bills that are wending their way through the legislative process.

Several states have enacted AI legislation that focuses specifically on data privacy and accountability. The law firm [BLCP keeps a running total of state-by-state legislation](#) updated quarterly. The US Chamber of Commerce offers its own [AI Legislation Tracker](#).

As of December 2024, 17 states have adopted laws regulating how AI can be used.



Examples include:

- ▶ The [California AI Transparency Act](#) (SB942) requires businesses to disclose AI-generated content and to “to make available an artificial intelligence (AI) detection tool at no cost to the user that meets certain criteria, including that the AI detection tool is publicly accessible.”
- ▶ Colorado enacted [Consumer Protections for AI](#) (SB24-205), which requires developers of companies that deploy high-risk AI systems to use reasonable care to avoid algorithmic discrimination.

Companies that prioritize Responsible AI are better positioned to meet these current and emerging legal standards, enabling them to avoid fines, lawsuits, or restrictions.

The EU AI Act

The EU AI Act will be fully implemented by 2025 and comes with significant legal requirements, including:

- Mandatory transparency about AI systems’ capabilities and limitations
- Risk-based classification system with stricter rules for high-risk AI
- Required human oversight for certain AI applications
- Disclosure when content is AI-generated
- Fines up to 7% of global revenue for violations

Companies that prioritize Responsible AI are better positioned to meet these current and emerging legal standards.

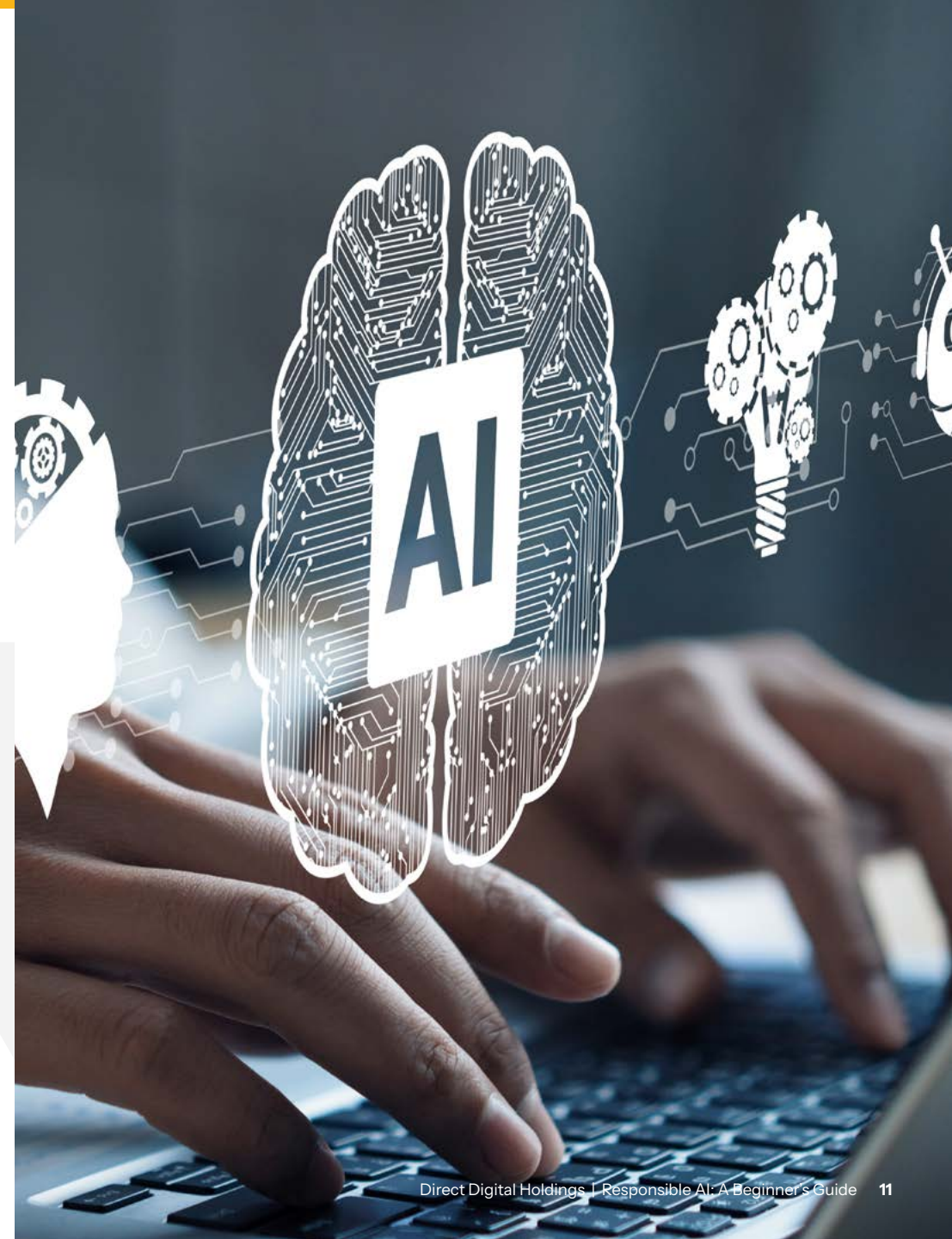
Why worry about the EU AI Act if your users are not EU citizens? It's important to note the "Brussels Effect," which we witnessed when the EU GDPR went into effect. Once the EU began regulating tracking, businesses across the globe voluntarily followed suit. The [Brussels Effect](#) explains why you encounter cookie banners requiring consent, even if you live in a state with no consumer data privacy regulations.

Improved Decision-Making

Improved decision-making is another key reason for implementing and following a Responsible AI framework. AI systems built responsibly are less likely to produce biased or inaccurate results. Ensuring that your AI outcomes are fair and transparent will lead to better decision-making, directly benefiting your organization's operations and strategy.

Risk Mitigation

Irresponsible use of AI can lead to significant risks, such as cybersecurity breaches, biased decision-making, and operational inefficiencies. By embedding Responsible AI principles into the design of your data, model design, and feedback loop (more on that later), you will proactively address these risks before they escalate into major problems.





Risk-Based Categorization of AI Systems

The legislation discussed in the previous section requires companies to take steps to eliminate risks from their AI tools, but you may be wondering: How is “risk” defined? Good question!

Most regulations and frameworks developed by AI leaders rely on risk-based categorizations of AI systems. Those categorizations are:

Minimal Risk	These tools, such as AI-powered spell checkers, are considered low risks because the consequences are minor if they make a mistake. In general, they don't require a lot of heavy monitoring.
Moderate Risk	These AI systems can cause significant but non-life-threatening harm. These tools make decisions, such as who to hire, who to approve for a car loan, and things like predictive infrastructure maintenance. You need to take active steps to eliminate bias and ensure accuracy and accountability.
High Risk	These AI systems carry significant risks (think AI for medical diagnoses or self-driving cars). Mistakes made by these tools can be life-threatening. For this reason, high-risk AI systems require strict safeguards, testing, and accountability.

Most regulations and frameworks developed by AI leaders rely on risk-based categorizations of AI systems.

The EU AI Act Risk Framework (see below) has a fourth risk level: Unacceptable.

Know Your Risk Level

You may assume that if your organization uses off-the-shelf tools, such as ChatGPT, that your risk level is low. But suppose your employees use an LLM-based tool for a variety of use cases, but that may not be the case. If, unbeknownst to you, employees are using a free version of the tool for client data, your organization is in jeopardy of a data leak.

You will know if you need a robust set of guardrails if your organization:

- Deploys AI that affects key decisions such as hiring, lending, resource allocation
- Uses AI across multiple business functions
- Handles sensitive or personal data, such as healthcare
- Serves vulnerable populations
- Deploys an AI system that affects many users

If your AI meets any of the above, you will need to adopt a Responsible AI framework to protect its users and your brand reputation.

Unacceptable risk (i.e., use case is banned)	<ul style="list-style-type: none">• Social scoring systems• Subliminal manipulation• Real-time biometric identification in public spaces• Emotion recognition in workplaces/schools
High Risk	<ul style="list-style-type: none">• Critical infrastructure• Education/vocational training• Employment/worker management• Access to essential services• Law enforcement• Migration and border control• Administration of justice
Limited Risk	<ul style="list-style-type: none">• Chatbots• AI-generated content• Emotion recognition systems• Biometric categorization
Minimal Risk	<ul style="list-style-type: none">• AI-enabled video games• Spam filters• Inventory management• Basic business tools

More Reading

The risk-based classification is a common approach across many AI governance frameworks. If you want to see additional risk-based approaches, check out the [NIST AI Risk Management Framework](#) or the [ISO/IEC standards use risk levels](#).



Key Pillars of Responsible AI

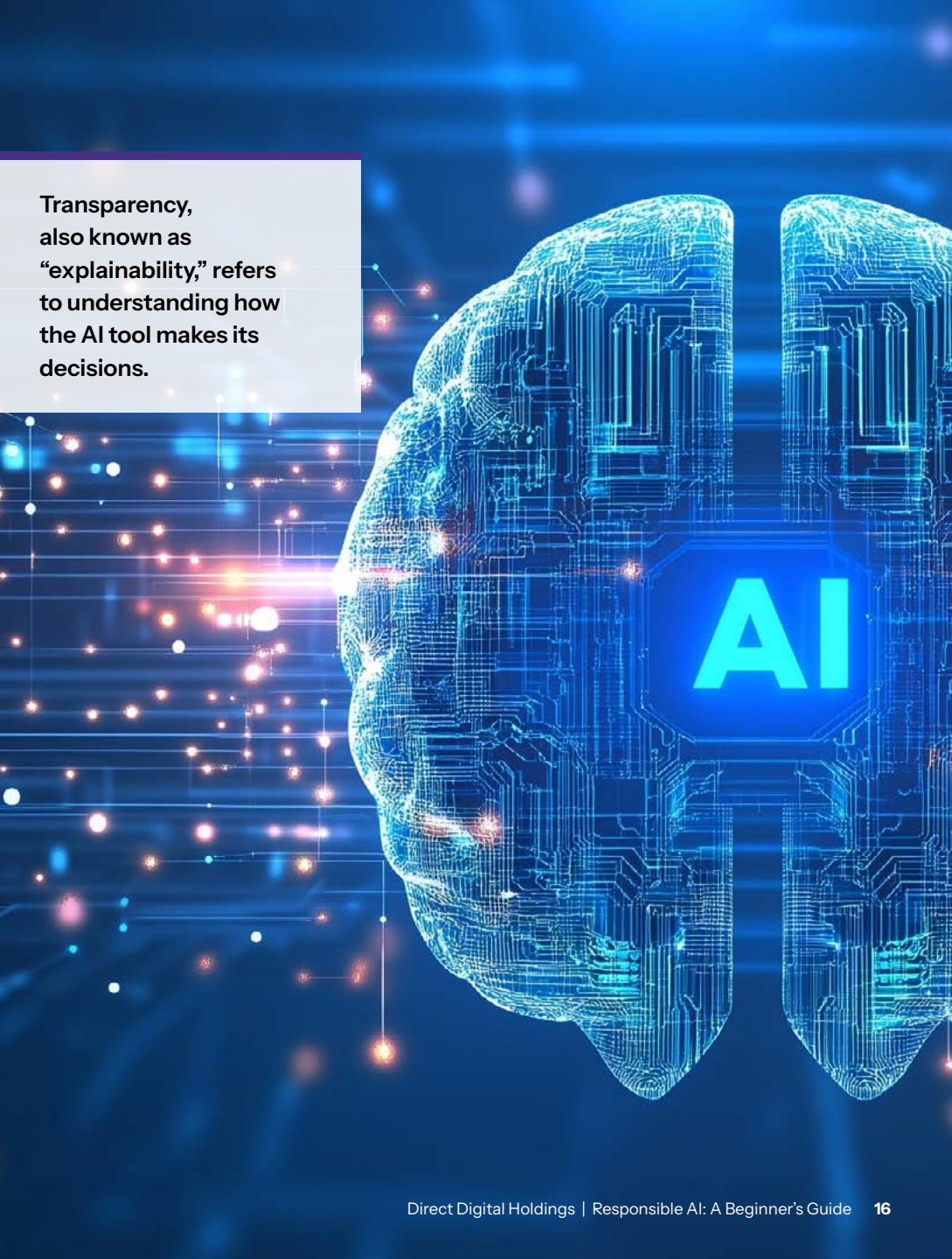


There are many Responsible AI frameworks available for you to reference, many developed by leaders in the industry, including Microsoft, HubSpot, IBM, NIST and others. You don't need to reinvent the wheel.

Some use slightly different terms to describe their essence, but in general, all incorporate the following key principles, generally referred to as pillars.

- ▶ **Transparency.** Transparency, also known as “[explainability](#),” refers to understanding how the AI tool makes its decisions. Can your organization explain how your model the parameters used to arrive at specific outcomes? When decisions are made without explanation, users and regulators distrust them, as they have the potential for misuse.
- ▶ **Fairness & Bias Mitigation.** Does your AI treat all groups equitably? Bias is rampant in AI because the data it is trained on can be biased (consider predictive policing, which is trained on inherently biased arrest records).

AI must treat all users equitably, especially in moderate and high-risk categories such as healthcare, hiring, or lending. If an AI system is biased, it can unfairly harm large groups of people.



Transparency, also known as “[explainability](#),” refers to understanding how the AI tool makes its decisions.

- ▶ **Accountability.** Accountability demands a person or team with the organization who will take responsibility for your AI tool. For instance, who is accountable for the AI's output? And who is responsible for addressing mistakes and addressing any harmful outputs that may occur?

Air Canada's chatbot provided false information and was sued. The company lost, in part, because there was no clear accountability or process for resolution.

Many companies build teams of AI experts who oversee their AI tools. Interestingly, job postings for AI ethics roles have increased by 106% since 2019. The U.S. Bureau of Labor Statistics projects AI Ethics Researcher jobs will grow by 28% from 2019 to 2029.

- ▶ **Privacy.** Consumer data protection is a cornerstone of every Responsible AI framework. All AI systems must respect user data, take steps to ensure it is secure, and use it only in ways that avoid breaches or misuse.
- ▶ **Reliability.** Reliability refers to the consistency of outputs. Does the model perform as intended? Does it deliver accurate results, and maintain stable performance over time? Reliability requires robust error handling, clear performance metrics, and regular testing to ensure dependable operation.



Existing Responsible AI Frameworks



As your organization moves into the run-and-fly stages of AI adoption, your requirements for Responsible AI will increase.

There are many frameworks available for you to follow.

Microsoft's Responsible AI Standard	<p>This framework is built upon six core principles: fairness, reliability and safety, privacy and security, inclusiveness, transparency, and accountability. It provides guidelines for building AI systems that align with these principles.</p>
HubSpot's Ethical Approach to AI	<p>HubSpot's AI framework is built upon core principles that guide AI systems' ethical development and deployment. These principles include transparency, security, and accountability. By adhering to these guidelines, HubSpot ensures its AI technologies are developed and utilized responsibly, aligning with its commitment to ethical practices.</p>
Google AI Principles	<p>Google has outlined a set of AI principles that serve as a framework for responsible development and use of AI. These principles emphasize objectives for AI applications and specify areas the company will not pursue, ensuring AI technologies are developed ethically.</p>
IBM's Principles for Trust and Transparency	<p>IBM's framework focuses on responsible AI by emphasizing trust and transparency. It provides guidelines for the ethical development and deployment of AI systems, ensuring they are designed to be fair, explainable, and secure.</p>
OECD AI Principles	<p>The Organisation for Economic Co-operation and Development (OECD) developed a global framework to promote innovative and trustworthy AI. It is built on inclusive growth, fairness, transparency, security, and accountability. These guidelines serve as a global benchmark and have been adopted by governments worldwide.</p>
NIST AI Risk Management Framework	<p>The National Institute of Standards and Technology (NIST) created a risk management framework to ensure AI systems are trustworthy, fair, and safe. It identifies, manages, and mitigates risks throughout the AI lifecycle, providing a structured approach to responsible AI governance.</p>
Partnership on AI (PAI) Framework	<p>The Partnership on AI (PAI) is a nonprofit coalition offering guidelines for ethical AI use in media, labor, and public health. It emphasizes fairness, transparency, and mitigating unintended consequences. Contributors include major organizations like Apple, Facebook, and Amazon.</p>





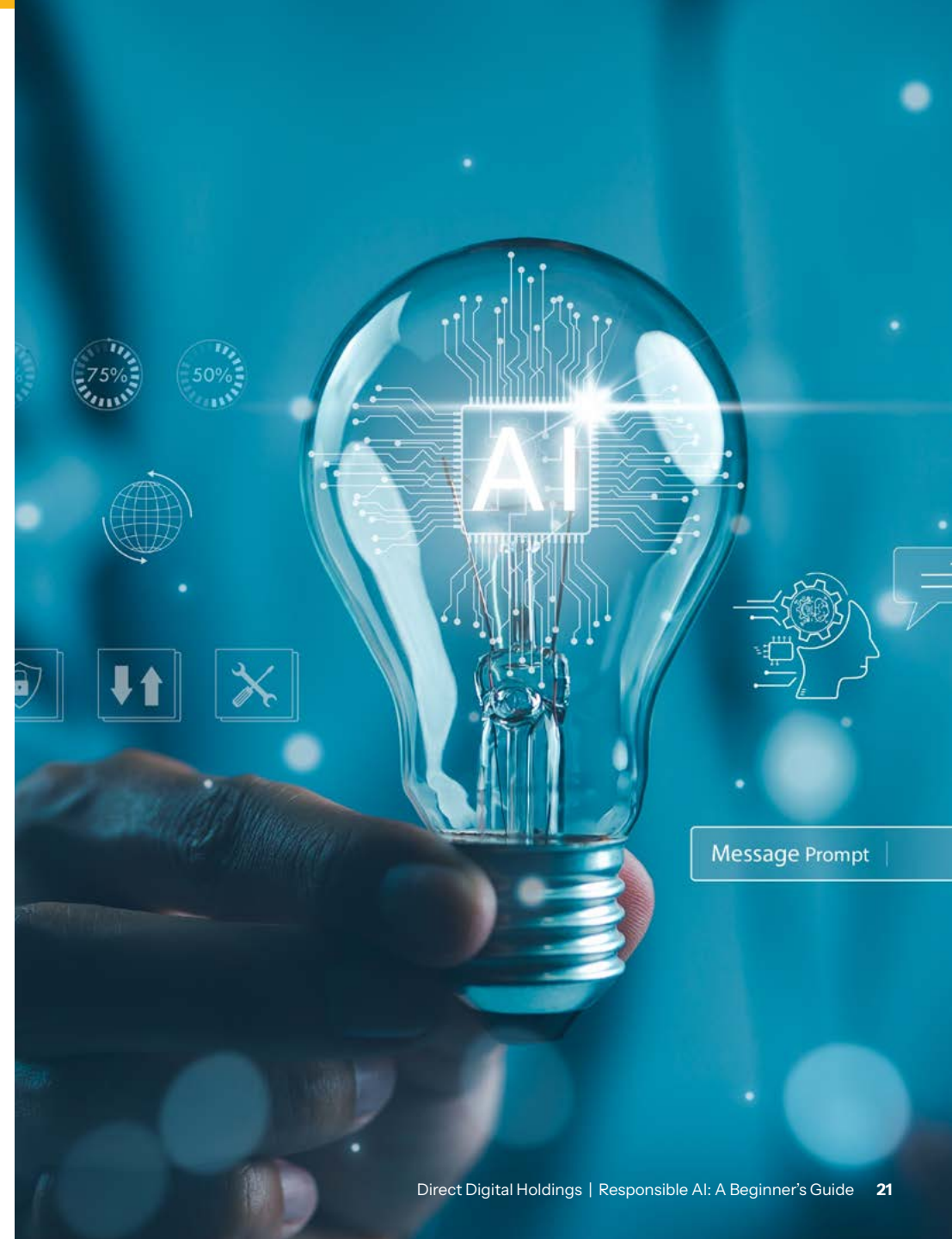
Responsible AI Governance



Let's take the basic pillars of Responsible AI and see how they apply to the core components of AI development: data, model design, and feedback loop.

Transparency: Goal: Build trust by ensuring AI decisions are understandable to users and stakeholders.

Data Questions	<ul style="list-style-type: none">• Can data sources and their transformations be documented for stakeholders?• Have you ensured transparency in the data-labeling process and potential biases introduced during labeling?
Model Design Questions	<ul style="list-style-type: none">• Can the AI system provide reasons for its outputs (Explainability)• Is there a mechanism for users to request clarifications about the decisions made by the AI?• Can you track and document model changes over time?
Feedback Loop Questions	<ul style="list-style-type: none">• Is there a mechanism for logging and reviewing user questions about decisions?• Who ensures that explanations are clear and accessible for non-technical audiences?



Fairness & Bias Mitigation: Goal: Ensure AI systems do not amplify bias and treat all groups equitably.

Data Questions	<ul style="list-style-type: none"> • Have you conducted bias audits on the data? • Are there systems to monitor for emergent bias as new data is added?
Model Design Questions	<ul style="list-style-type: none"> • Have you stress-tested (i.e. red teamed) the model under diverse scenarios to identify bias? • Can your system flag and adjust for outliers or patterns of unfair treatment?
Feedback Loop Questions	<ul style="list-style-type: none"> • Is there a reporting mechanism for bias-related complaints? • How frequently is model performance reviewed for fairness across demographic groups?

Accountability: Goal: Ensure AI systems comply with regulations, have oversight, and are auditable.

Data Questions	<ul style="list-style-type: none"> • Is there a designated owner for each dataset to ensure data quality and compliance? • Are you logging who accesses and modifies the data?
Model Design Questions	<ul style="list-style-type: none"> • Can the system's decision-making process be audited? • Have you established clear roles for governance, such as a Responsible AI officer or board?
Feedback Loop Questions	<ul style="list-style-type: none"> • Are there documented response plans for addressing failures or harmful outcomes? • Is there an escalation process for unresolved issues?

Privacy: Goal: Protect user data and ensure ethical data use.

Data Questions	<ul style="list-style-type: none"> • Are you using privacy-preserving techniques like differential privacy or federated learning? • Have you verified compliance with jurisdiction-specific privacy laws such as GDPR or CCPA?
Model Design Questions	<ul style="list-style-type: none"> • Have you applied data minimization principles (only collecting and using what is necessary)? • Can users control how their data is used (e.g., opt-out mechanisms)?
Feedback Loop Questions	<ul style="list-style-type: none"> • Is there a process for users to request deletion or correction of their data? • Are regular privacy impact assessments conducted?

Reliability: Goal: Ensure models perform as expected under real-world conditions.

Data Questions	<ul style="list-style-type: none"> • Do you have the right data to train your model to be accurate? • Is the data properly labeled? • Are data validation and version control systems in place to ensure consistency? • Is synthetic data used when real data is insufficient, and is it validated for quality?
Model Design Questions	<ul style="list-style-type: none"> • Do you have the right model to meet your AI goal? • Have you proto-typed your model? • How will you move it into production in a cost-effective and accurate way? • What are your precision/recall goals?
Feedback Loop Questions	<ul style="list-style-type: none"> • Have you established human oversight to monitor outputs? • How often are reviews conducted? • Have you established response types by level of the problem?

Parting Thoughts: Integration Across Your Governance Framework



Parting Thoughts: Integration Across Your Governance Framework

Building responsible AI takes more than just having rules on paper — it requires real action and follow-through by specific people within your organization. Those actions include getting input from users, measuring what matters, and having teams work together to keep improving how AI systems work.

The key pieces — being open about how AI works, making sure it treats everyone fairly, taking responsibility for outcomes, protecting privacy, and building reliable systems — need to be an integral part of how your organization operates. When you get this right, you build trust and create AI that works for everyone.

Want to Learn More?

Check out our additional AI resources:

- ▶ **Demystifying Generative AI**
- ▶ **The Generative AI Playbook: Implementation & Best Practices**
- ▶ **Responsible AI: A Beginner's Guide**
- ▶ **Responsible AI**

[Click Here](#)



Glossary of Terms

Artificial Intelligence (AI): The field of computer science focused on building systems capable of performing tasks that typically require human intelligence, such as decision-making, speech recognition, and language translation.

Generative AI: A subset of AI that creates new content—such as text, images, or code—based on patterns learned from existing data, often using models like GPT or DALL-E.

Responsible AI: A framework ensuring that AI tools are unbiased, safe, transparent, and accountable, minimizing risks to users and organizations while maintaining trust.

LLM-Based Tools (Large Language Model-Based Tools): Applications or systems built on large language models, such as GPT or Claude, that process and generate human-like text based on extensive training data.

Accountability: The principle that ensures clear ownership of AI tools, with designated individuals or teams responsible for outcomes, monitoring, and addressing any harmful results.

Bias: A systematic error in AI outputs that unfairly favors or disadvantages certain groups, often stemming from biased training data or flawed algorithms.

Data Privacy: The practice of protecting sensitive user information, ensuring it is used ethically and complies with privacy laws such as GDPR or CCPA.

Differential Privacy: A privacy-preserving technique that allows organizations to derive insights from data while minimizing the risk of identifying individual users.

Explainability (Transparency): The ability to understand and articulate how an AI system makes decisions, ensuring trust and enabling users to interpret its outputs.

Fairness: The principle that AI systems must treat all individuals equitably, without discrimination or bias, especially in critical areas like hiring, lending, and healthcare.

Federated Learning: A privacy-focused machine learning approach where algorithms are trained across decentralized devices or servers without transferring raw data to a central location.

Governance: A structured framework for managing and overseeing AI systems, ensuring they align with ethical, legal, and organizational standards.

High-Risk AI: AI systems that pose significant potential harm, such as those used in healthcare, law enforcement, or autonomous vehicles. These require strict safeguards, testing, and accountability measures.

IBM's Principles for Trust and Transparency: A framework focused on ensuring that AI systems are fair, explainable, and secure, emphasizing trust and transparency in AI deployment.

Minimal Risk AI: AI systems with low potential for harm, such as spell checkers or spam filters, which require minimal monitoring and oversight.

Moderate Risk AI: AI systems that can cause significant but non-life-threatening harm, such as tools for hiring or loan approvals. These require bias mitigation and accountability.

NIST AI Risk Management Framework: A framework from the National Institute of Standards and Technology that provides guidance for identifying, assessing, and mitigating risks in AI systems to ensure they are trustworthy, fair, and safe.

OECD AI Principles: Guidelines from the Organisation for Economic Co-operation and Development that promote inclusive growth, transparency, and accountability in AI systems.

Privacy-Preserving Techniques: Methods like differential privacy and federated learning that protect sensitive data while enabling insights or model training.

Risk-Based Categorization: A method of classifying AI systems based on their potential harm, ranging from minimal to high risk, with corresponding safeguards.

Transparency: The principle of providing clear and understandable explanations of how AI systems operate, fostering trust and accountability.

Unacceptable Risk AI: AI systems banned under regulations like the EU AI Act due to their potential for severe harm, such as social scoring or subliminal manipulation.

Microsoft's Responsible AI Standard: A framework built on six principles: fairness, reliability, safety, privacy, inclusiveness, transparency, and accountability. It offers guidelines for ethical AI use.

HubSpot's Ethical Approach to AI: A framework emphasizing transparency, security, and accountability in AI, ensuring tools are developed and deployed responsibly.

Google AI Principles: A framework that outlines objectives for ethical AI applications, emphasizing fairness and transparency while avoiding harmful or unethical pursuits.

Partnership on AI (PAI) Framework: Guidelines from a nonprofit coalition for ethical AI use in areas such as media, labor, and public health, emphasizing fairness and mitigating unintended consequences.

Disclaimer: The responses provided by this artificial intelligence system are generated by artificial intelligence based on patterns in data and programming. While efforts are made to ensure accuracy and relevance, the information may not always reflect the latest data and programming news or developments. This artificial intelligence system does not possess human judgment, intuition, or emotions and is intended to assist with general inquiries and tasks. Always conduct your own independent in-depth investigation and analysis of ANY information provided herein, and verify critical information from trusted sources before making decisions.

Interested parties should not construe the contents of ANY responses and INFORMATION PROVIDED herein as legal, tax, investment or other professional advice. In all cases, interested parties must conduct their own independent in-depth investigation and analysis of ANY responses and information provided herein. In addition, such interested party should make its own inquiries and consult its advisors as to the accuracy of any materials, responses and information provided herein, and as to legal, tax, and related matters, and must rely on their own examination including the merits and risk involved with respect to such materials, responses and information.

We nor any of our affiliates or representatives make, and we expressly disclaim, any representation or warranty (expressed or implied) as to the accuracy or completeness of the materials, responses and information PROVIDED or any other written or oral communication transmitted or made available with respect to such materials, responses and information or communication, and we, nor any of our affiliates or representatives shall have, and we expressly disclaim, any and all liability for, or based in whole or in part on, such materials, responses and information or other written or oral communication (including without limitation any expressed or implied representations), errors therein, or omissions therefrom.



 Colossus SSP*  Orange 142*

For more information:

Direct Digital Holdings
1177 West Loop South | Suite 1310
Houston, TX 77027
marketing@directdigitalholdings.com

Digital advertising built for everyone.

About Direct Digital Holdings

Direct Digital Holdings (Nasdaq: DRCT) brings state-of-the-art sell- and buy-side advertising platforms together under one umbrella company. Direct Digital Holdings' sell-side platform, Colossus SSP, offers advertisers of all sizes extensive reach within the general market and multicultural media properties.

The Company's buy-side platform, Orange 142, delivers significant ROI for middle-market advertisers by providing data-optimized programmatic solutions for businesses in sectors ranging from energy to healthcare to travel to financial services. Direct Digital Holdings' sell- and buy-side solutions generate billions of impressions per month across display, CTV, in-app, and other media channels.

To learn more please visit directdigitalholdings.com

