



Advancing Trusted AI in ArcGIS

Trusted Artificial Intelligence (AI) goes beyond simply achieving accurate results for ArcGIS products. It is a journey of continuous advancement that encompasses a holistic approach prioritizing security, privacy, transparency, fairness, reliability and responsible development and deployment of AI. We recognize the power of AI technology and its potential to transform society and design a better, more sustainable future. The rapid advancements in generative AI highlight the urgent need for frameworks that guide trusted AI deployment and bridge the AI trust gap.

Legacy of AI Trust and Innovation

Before the generative AI boom of the past few years, when people talked about AI, typically they were referring to machine learning and deep learning models used in pattern recognition, forecasting, object detection, change detection, and more. Over a decade ago, Esri started with machine learning to perform clustering, regression, and classification on spatial data. More recently, work has continued in both the machine learning and deep learning space, including the introduction of [pre-trained deep learning models](#) that make it easier to get started with tasks such as feature extraction, point cloud classification, and image redaction. This work falls under the umbrella of what Esri refers to as [GeoAI](#).

Generative AI refers to a type of machine learning model designed to create new content, or insightful recommendations, by learning patterns from large-scale datasets. Generative AI models are trained on extensive datasets, which may include very large, domain-specific data and/or data from the internet. Unlike traditional AI models that focus on predictive analytics, generative AI models are used to generate creative outputs such as text, images, or other forms of content.

In contrast to Generative AI as described above, GeoAI focuses on analyzing and interpreting geospatial data to uncover patterns and perform predictive analytics using the input data. Generative AI can be thought of as a machine learning model that is trained to create new data, rather than making a prediction about a specific dataset and is more non-deterministic allowing for more creative solutions. A generative AI system is one that learns to generate more objects that look like the data it was trained on. An example of ArcGIS capabilities incorporating generative AI are [AI Assistants within ArcGIS](#).

The actual machinery underlying generative AI and other types of AI oftentimes utilize the same algorithms, which can blur the distinction between the types. Generative AI's quick proliferation and broader use cases has resulted in expedited regulatory requirements and customer demands for stronger transparency and control. Therefore, this paper primarily focuses on assurance measures being worked on or already in place for generative AI.

AI Landscape Today

The AI landscape is rapidly evolving. Governments around the world are actively shaping the future of AI by enacting new laws and frameworks. For instance, the European Union recently adopted the EU AI Act, which establishes regulations on high-risk AI applications. The United States has introduced various AI governance initiatives, including the AI Bill of Rights and executive actions, to address potential risks, biases, and safety concerns in AI, though these policies continue to evolve.

Esri has recognized the importance of staying ahead of these evolving legal and regulatory requirements, responsible development, and ethical considerations. We proactively align our AI practices with key regulations and industry-recognized frameworks. This includes following the guidance set forth by laws and regulations stated above.

ArcGIS Guiding AI Principles

Esri's dedication to trusted AI is rooted in our core values, driving us to innovate with integrity. Our AI Principles guide our AI development and deployments, helping ensure our systems positively impact society, provide transparency, and protect user data. Esri's Trusted AI is anchored on a foundation of six core principles that guide our AI projects and initiatives.



Security: We are committed to safeguarding security and mitigating risks in our AI systems through a secure-by-design approach while ensuring responsible AI that proactively protects against security threats.



Privacy: We prioritize protecting user data and ensuring the privacy of AI throughout the AI lifecycle ensuring compliance with global privacy standards through privacy-by-design methodologies, data anonymization, and data minimization.



Transparency: We provide clear visibility about our AI models, empowering informed decision-making about our AI processes, limitations, and outcomes.



Fairness: Esri has long upheld the principles of fairness, ethics, and societal responsibility in its everyday practices. These core values are embedded in our approach to decision-making, product development, and community engagement.



Reliability: Our AI is carefully tested and validated to deliver consistent and dependable results across diverse environments and use cases.



Accountability: We maintain accountability by establishing clear governance frameworks, holding our teams responsible for AI deployment and monitoring, ensuring human oversight remains central to all AI-related decisions.

Esri's Approach to Trusted AI

We have implemented a variety of measures and practices to ensure our guiding AI principles are translated into action.

Design

Our approach to AI development at Esri is grounded in practices designed to build and maintain trust. We employ a risk evaluation process to assess new AI products and features, ensuring they meet relevant privacy and security standards. Adhering to a human-in-the-loop design philosophy, our AI features are developed to support inclusivity and user control. We are establishing red teaming and poisoned model validation techniques to identify vulnerabilities and potential biases by simulating adversarial attacks and validating against manipulated data. We ensure incorporation of ethical guardrails to ensure compliance with ethical principles. Our generative AI solutions also undergo holistic lab testing in controlled environments, coupled with our human oversight framework to maintain human involvement in critical decision-making processes supported by AI.

Customer Choice

Esri prioritizes transparency and user control when it comes to Generative AI by informing customers when Generative AI is being used and providing them with alternative options. Due to the broader societal concerns with some use cases, or other trust concerns, some organizations choose to block generative AI. With Esri products, generative AI capabilities are only enabled when customers opt-in, ensuring that users have full autonomy over their use. For customer managed products, this can include optional installation components and administrative override settings for services.

Data

Esri prioritizes your data security and control with our products including when utilizing AI functionalities within ArcGIS. We understand the importance of Trusted AI, and that includes transparency in data handling. Unless explicitly authorized by you (such as providing feedback within the application), Esri does not use customer prompts to train AI models. Customers retain ownership of prompts and data they provide for AI analysis within ArcGIS products.

In the instances where we leverage a third-party AI service for specific AI functionalities, we ensure they adhere to data segmentation practices. Esri only collaborates with external services that commit to robust data handling practices, including Esri's use of enterprise-class AI instances that segment and protect data. By prioritizing these practices, we ensure your data used with AI Assistants is not sold or used for training Esri or third-party owned AI models unless the customer authorizes Esri to do so. Note, to ensure a safe experience, most external service providers temporarily store anonymized prompt information via tokenization and hashing for abuse monitoring purposes.

Metadata

The importance of metadata continues to increase relative to AI. Whether it is metadata concerning the underlying model, or dataset licensing which will determine if the information can be processed by an AI solution. While many underlying AI models utilized by our products have AI Model Cards (an evolving standard), they don't provide the additional context for how the model is utilized within our product and what steps Esri has taken to minimize associated risks. Therefore, Esri is creating [AI Transparency Cards for ArcGIS](#) features that utilize generative AI. These cards provide details on feature functionality, data sources, validation performed, and safeguards in place allowing our customers to incorporate and utilize such features in a responsible manner.

Governance

Esri established a cross-department AI Governance Board in early 2023 to establish policies and procedures to ensure adherence to our Trusted AI principles. This fosters a culture of collaboration and innovation, allowing us to leverage the power of AI while prioritizing trust and responsibility for both our products we deliver to customers, as well as our internal operations.

Moving Forward with Trust

Note that achieving Trusted AI is a shared responsibility with our customers. While we strive to ensure trusted AI tools, their full potential can only be realized through active customer participation. This includes adhering to best practices in data governance, ensuring proper usage and handling of AI tools, providing feedback, ensuring appropriate human oversight, and remaining aware about the ethical implications of AI deployment. By doing so, customers contribute to the integrity and trustworthiness of Esri AI applications, fostering a collaborative environment where both Esri and its users can innovate responsibly and sustainably.

Our journey with AI has been one of continuous evolution, from the powerful foundations of GeoAI to the exciting possibilities of Generative AI Assistants. Through this journey, our dedication to trust has remained constant. The principles and practical initiatives serve as a roadmap for our continued commitment to our AI solutions that are not only influential but also reliable, ethical, and transparent.

Together, we can build a future where AI empowers positive change with trust as our foundation.

Latest document version available within the [ArcGIS Trust Center here](#).