



The Deloitte On Cloud Podcast

Title: Building trust in AI: Deloitte's Ganesh Seetharaman unpacks trustworthy, accountable AI

Description: In this Knowledge Short, Ganesh Seetharaman delves into trustworthy AI and why it's foundational, not optional. He explains how observability can uncover bias, ensure fairness, and foster transparency. Ganesh also discusses frameworks and oversight tools organizations can use, such as real-time monitoring and unified dashboards. These approaches improve governance, reduce operational risk, and support responsible deployment at-scale. Finally, he explores the road ahead for trustworthy AI.

Duration: 00:07:43

Ganesh:

Welcome to this Deloitte On Cloud Podcast Knowledge Short. I'm Ganesh Seetharaman, managing director within Deloitte's AI, Data, and Engineering practice. Today, we are engaging a question at the very core of digital transformation. How do leaders build AI that organizations and public can trust? In this episode, I'll unpack how platforms, coupled with advances in observability, are shaping not just out of the possible, but also defining the frameworks required to govern AI with clarity and accountability.

Let's begin with the reality facing every executive. Trust in AI isn't just a differentiator. It's a foundational demand. From financial underwriting and supply chain optimization to clinical decision support, AI agents today can impact real lives and business outcomes. In these moments, opacity doesn't simply hinder process or progress, it introduces systematic risk for your brand and businesses. AI trustworthiness isn't merely about avoiding technical mishaps. We must confront deeper issues like unconscious bias, embedded data, opaque decision pipelines, and outcomes that elude explanation. When transparency falters, stakeholder confidence erodes and regulatory risk mounts. Trustworthy AI should be built on the principle that systems must be explainable, fair, and accountable throughout their life cycle.

Imagine deploying a sophisticated AI-driven virtual assistant in a leading hospital. This agent triages patient requests and guides care decisions where the stakes are regulatory-compliant, patient safety, and ultimately brand reputation. Every new model capability compounds both opportunity and risk. Can you demonstrate fairness in its choices, audit its behavior when challenges arise? For mission-critical applications, transparency and reliability are not nice-to-haves. They are non-negotiable.

Think of it as building a control tower for digital operations. Just as we would not let a pilot fly without instruments, AI must operate with robust oversight mechanisms. This is where observability comes into play. Think of observability as your digital instrumentation panel and dashboard for your AI agent.

Using the hospital example above, let's break down what modern observability AI really means. One: real-time tracking—every patient interaction is logged instantly, so you can monitor request types, urgency, wait times, and health threats. Two: audit trails—every recommendation like go to the ER or self-manage at home is saved along with the logic behind it. Auditors can retrace decisions down to the data and the AI model versions used. Three: the outlier alerts—if the chatbot makes an unusual or a risky suggestion, the system flags it, alerting clinical staff before any harm can occur. Four: bias and drift deduction—observability tools should scan thousands of interactions watching for patterns or shifts, especially as certain groups being treated differently or indifferently.

Is this AI device evolving inextricably? These are all caught early for investigation. Five: system health monitoring—the same dashboard ensures the healthy server resources or infrastructure resources that quickly resolves for any performance bottlenecks, visualize the relay rates as your agents talk to their other apps, flagging where slowdowns or errors occur or creep in. Six: continuous feedback loop—staff tag questionable responses, feeding real-world data back to the model for retraining. Over time, you will see the measurable improvements improving the model along the way.

Think of the AI model as an industrial engine handling thousands of prompts daily. The visibility that is required across every single part of interaction is key, and that's where the model toolkit is going to be built, and it's going to be multiple components like token monitoring. Instantly see how the model uses every unit of language it processes; outlier access alerts; get notified about odd behaviors, like a chatbot suddenly producing walls of text or consuming excess resources; scalable prompt experimentation, enabling the engineers to tweak and tune how the model responds, then watch those effects unfold across hundreds and thousands of use cases in real time.

What is the road ahead for unified, actionable, trustworthy AI? So, here's where the platforms and the observability vendors are all leading to. One, they're trying to build unified observability platforms. Soon, all the agents, models, and devices, be it a cloud or edge, will feed into a single transparent dashboard—LLM instrumentation. Large language models or small language models or any other neural networks, including the other agents, will be instrumented so that every decision is logged and contextually explained, so no more black boxes. Harmony across the systems, whether the AI operates on a factory floor or within the public cloud, audit and analytics will be unified and will be accessible and will be consumable. The smart rules will catch the bias, glitches, or runaway chaos before they become a crisis. And, most importantly, every trace and alert and insight feeds a continuous improvement loop, helping teams refine, retrain, and innovate with confidence.

To sum it up, modern AI observability isn't just about avoiding outages or addressing resilience or building a moral; it's about enabling innovation while keeping trust and transparency front and center. Remember, building trustworthy AI means combining advanced technology with timeless principles: observe, explain, and refine.

Thanks for tuning in. If you enjoyed this episode, be sure to subscribe for more deep dives into the tech shaping our world. Until next time, stay curious and stay informed.

Operator:

This podcast is produced by Deloitte. The views and opinions expressed by podcast speakers and guests are solely their own and do not reflect the opinions of Deloitte. This podcast provides general information only and is not intended to constitute advice or services of any kind. For additional information about Deloitte, go to Deloitte.com/about.

This publication contains general information only and Deloitte is not, by means of this publication, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This publication is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this publication.

Visit the On Cloud library
www.deloitte.com/us/cloud-podcast

About Deloitte

As used in this podcast, "Deloitte" means Deloitte Consulting LLP, a subsidiary of Deloitte LLP. Please see www.deloitte.com/us/about for a detailed description of our legal structure. Certain services may not be available to attest clients under the rules and regulations of public accounting. Please see www.deloitte.com/about to learn more about our global network of member firms. Copyright © 2025 Deloitte Development LLC. All rights reserved.