

From Black Box to Glass Box: Making AI Transparent

**Interact, Analyse, Understand:
Master Explainable AI**

Explainable AI (XAI)

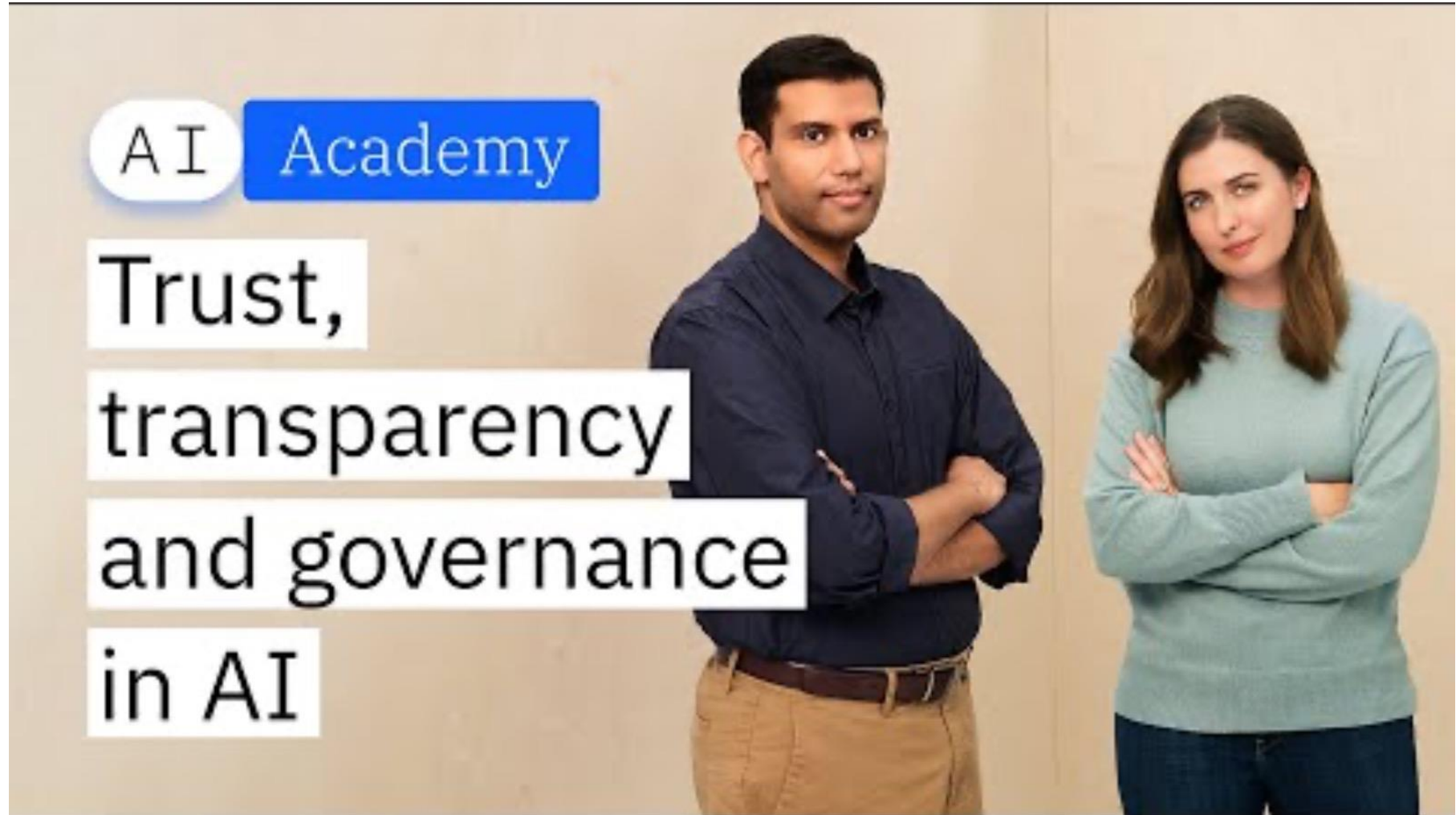


- ❖ **What is XAI?:** AI systems made clear and understandable to humans.
- ❖ **From Black Box to Glass Box:** Turns mysterious AI into something transparent by showing internal model attention.
- ❖ **Real-Life Example:** AI picking your poly course - wouldn't you want to know why?
- ❖ **Building Trust:** Essential for jobs like healthcare. Crucial for ethical AI and bias mitigation.
- ❖ **Singapore's Vision:** Fits AI Singapore's push for fair, safe tech everyone can trust. Mandatory for compliance.
- ❖ **Why It Matters to You:** Future employers want AI *expertise* – the ability to ***use and explain***. Make informed decisions.

Trust, transparency and governance

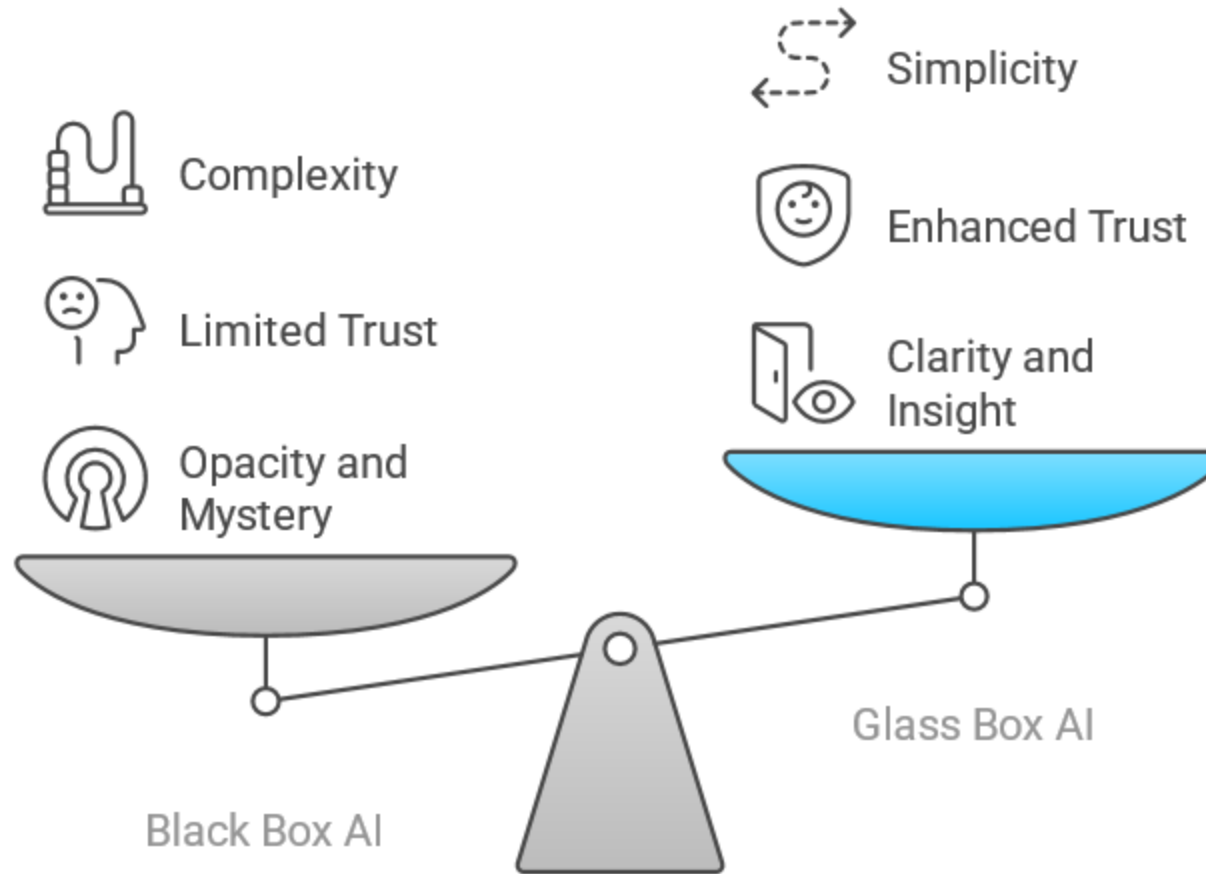


- ❖ Watch the following 9-minute video better understand why there is a need for explainable AI.



Ai

Transitioning from Opacity to Clarity in AI

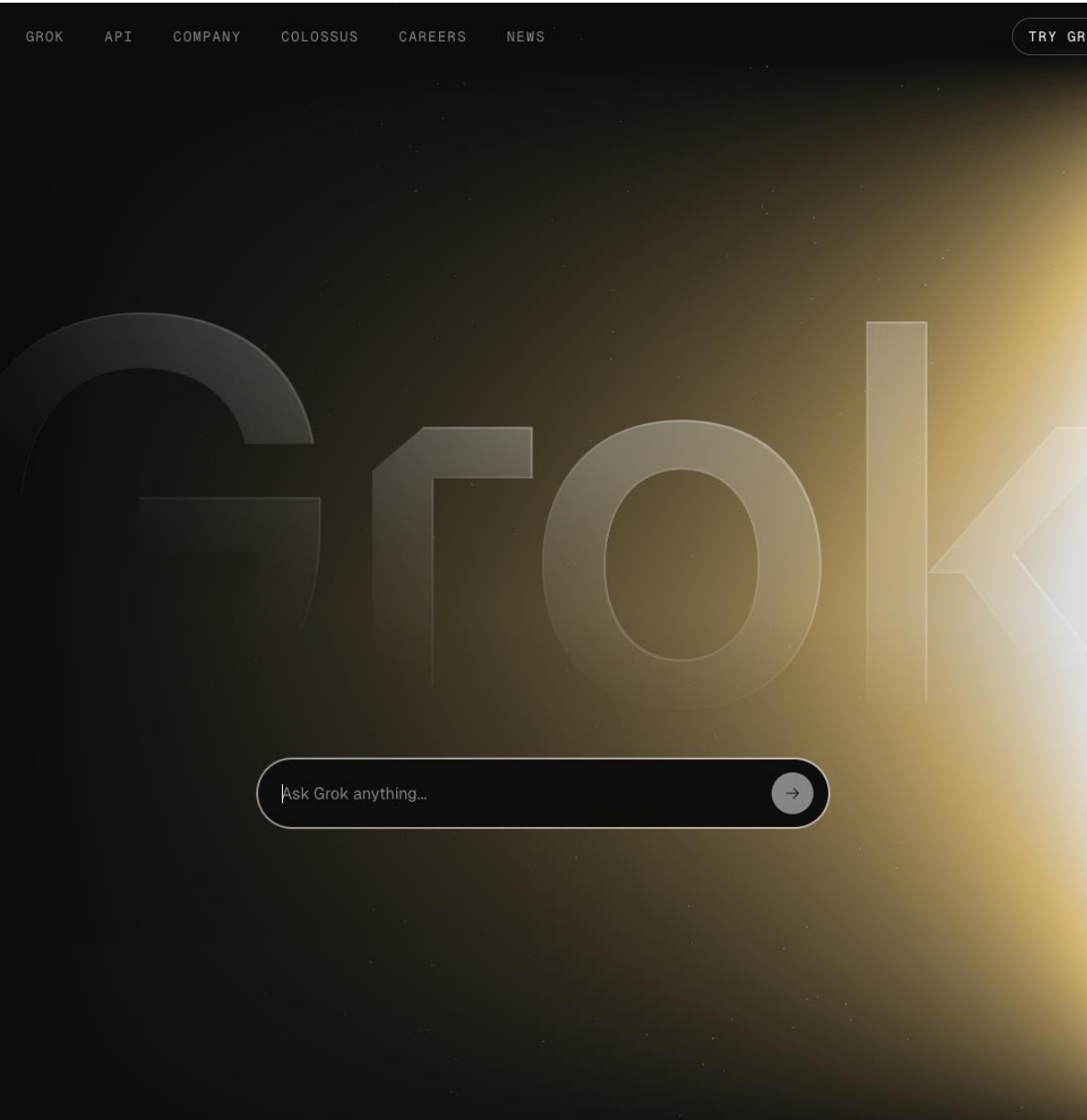




Grok

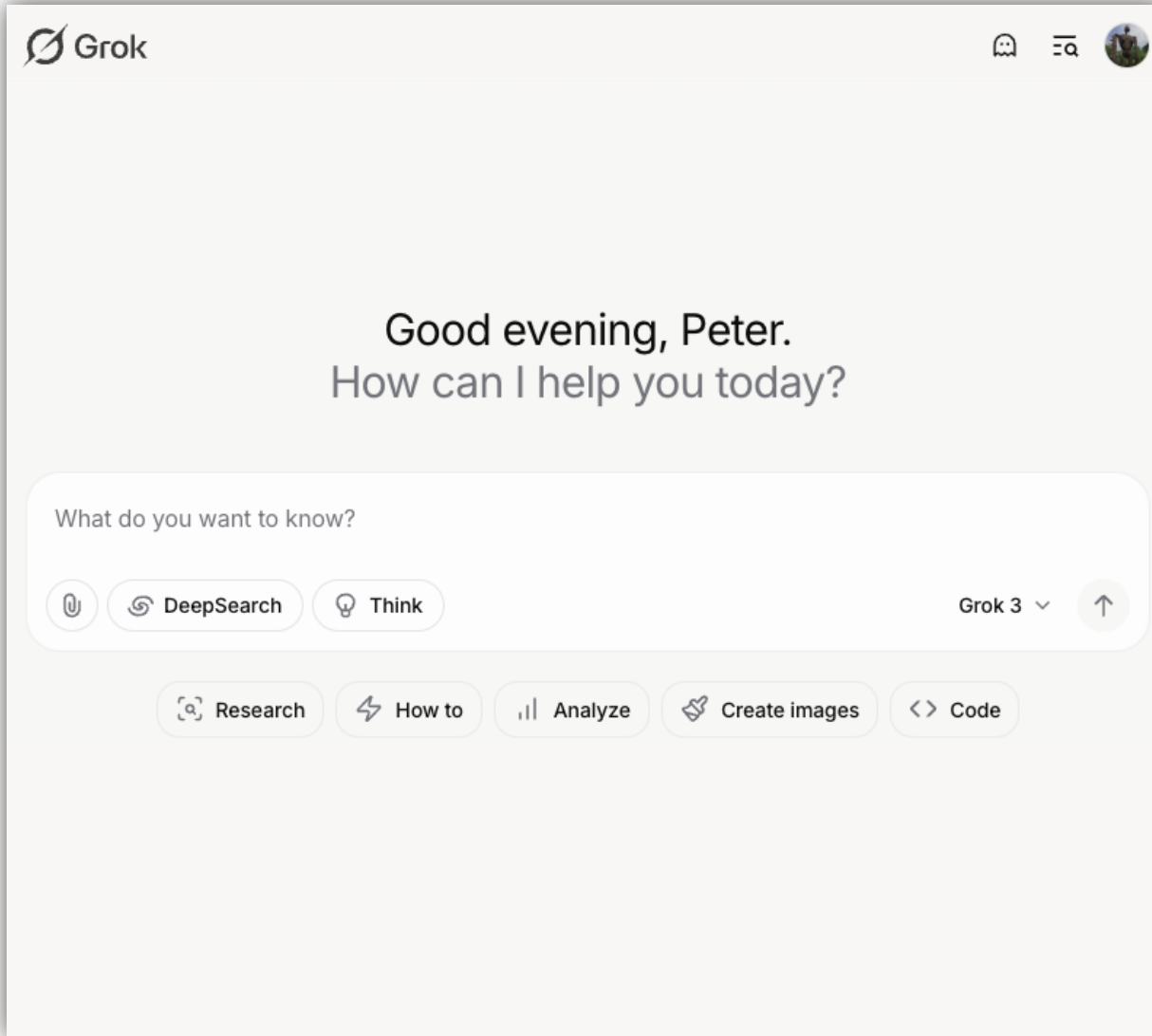
Self Description: *I'm Grok, xAI's cheeky AI pal who dishes out witty cosmic truths!*

Why use Grok to Investigate Explainable AI?



- ❖ **Free Access:** Available at no cost, perfect for students exploring XAI on a budget.
- ❖ **Interactive Responses:** Answers questions like “Why did you say that?” to show AI reasoning.
- ❖ **Real-Time Insights:** Pulls current X data, offering fresh examples for XAI analysis.
- ❖ **Simplified Explanations:** Breaks down complex AI ideas into clear, student-friendly terms.
- ❖ **Hands-On Learning:** Encourages active questioning to uncover how AI decisions work.
- ❖ **Unique Perspective:** Adds humour and wit, making XAI exploration engaging and fun.

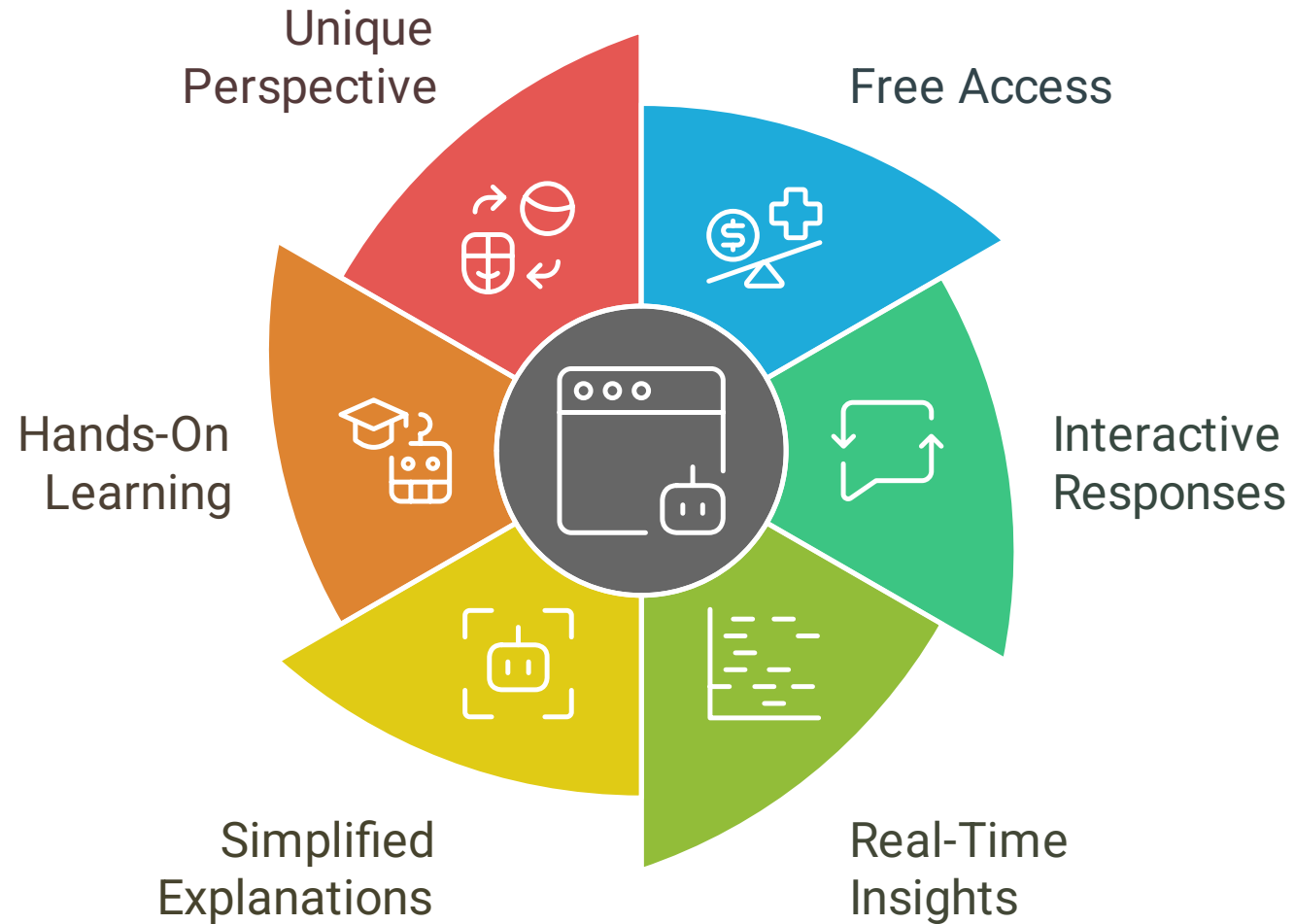
Grok 3 Free Tier: Capabilities



Sign in with X or Google (Recommended)

- ❖ **Conversational AI Access:** Useful: Ask Grok to explain its answers (e.g., “Why did you choose this?”) to peek into AI decision-making
- ❖ **Real-Time X Data:** Useful: Analyse current X posts to see how Grok interprets real-world data, revealing biases or reasoning
- ❖ **Basic Question Handling:** Useful: Pose “why” or “how” questions (e.g., “How did you decide?”) to uncover transparency limits
- ❖ **DeepSearch Lite:** Useful: Summaries of web content help compare Grok’s explanations to external sources
- ❖ **Think Mode Lite:** Useful: Step-by-step breakdowns (e.g., math problems) show interpretable AI processes
- ❖ **File Upload Analysis:** Useful: Upload sample data (e.g., text) and ask Grok to explain its analysis, testing clarity

Grok 3 Free Tier: Capabilities



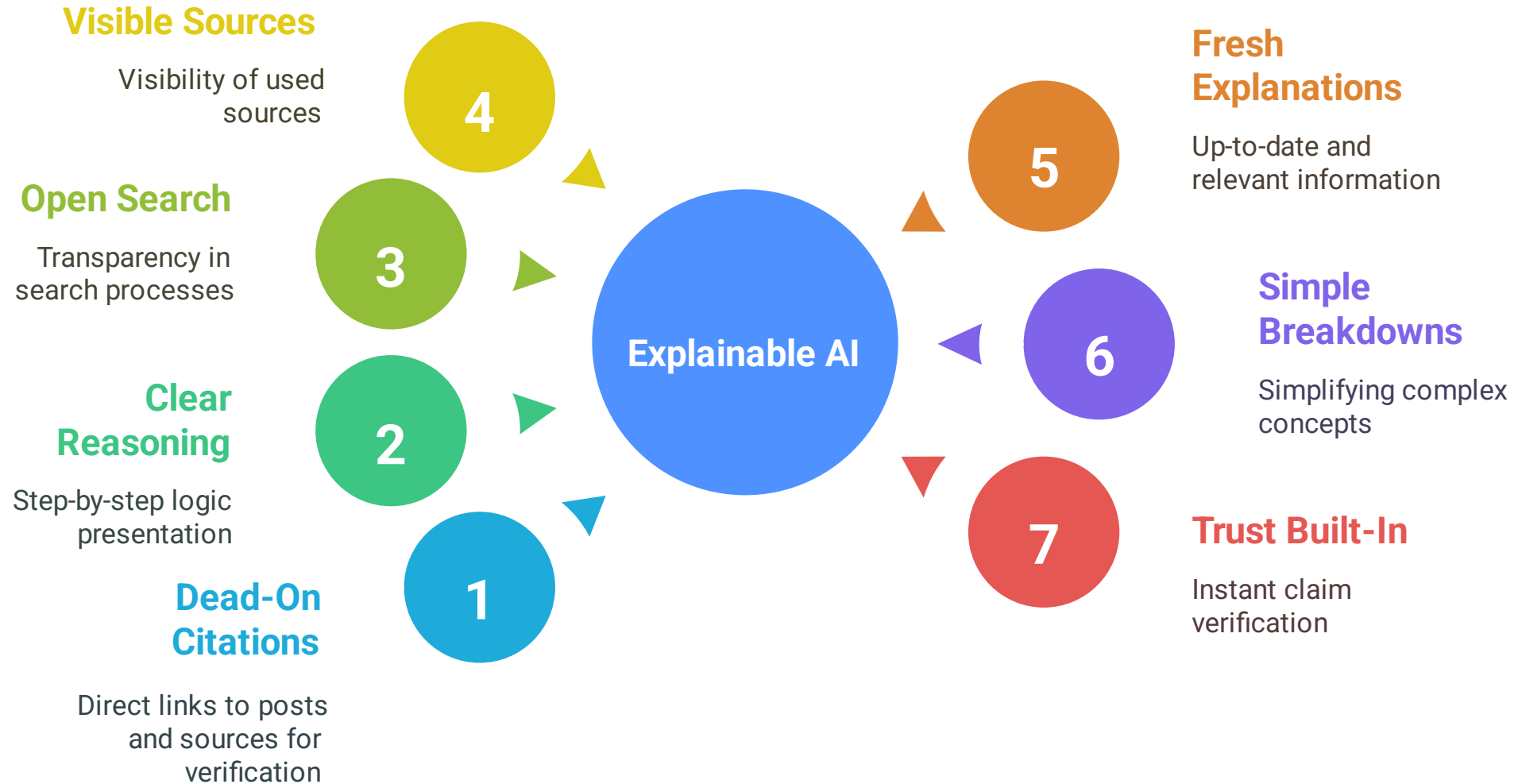
How Grok supports Explainable AI (XAI)



- ❖ **Dead-On Citations:** Links straight to X posts and web sources.
- ❖ **Clear Reasoning:** Shows its logic, step by step.
- ❖ **Open Search:** Reveals how it digs through X and the web.
- ❖ **Visible Sources:** See exactly what Grok uses.
- ❖ **Fresh Explanations:** Always up-to-date and relevant.
- ❖ **Simple Breakdowns:** Makes responses easy to grasp.
- ❖ **Trust Built-In:** Verify claims on the spot.



How Grok supports Explainable AI (XAI)



Trust, transparency and governance



- ❖ Watch the first 15 minutes of the video below to get an excellent overview of Grok and its various capabilities.



Ai



Explainable AI

Grok

Examples of XAI in Action:



❖ Healthcare:

- **Use:** AI diagnostics (e.g., tumour detection, anomaly detection in medical imaging).
- **Goal:** Doctors and patients understand why a tumour is flagged or an anomaly detected, leading to informed treatment decisions.

❖ Finance:

- **Use:** Credit risk assessment.
- **Goal:** Ensure fair lending by explaining approval/denial decisions, contributing to financial inclusion and transparency.

❖ Law:

- **Use:** Predicting case outcomes.
- **Goal:** Transparent legal decisions for judges and clients, facilitating understanding of the rationale behind court rulings.

❖ Education:

- **Use:** Personalized learning paths.
- **Goal:** Provide educators with deeper insights into student learning needs, improve learning outcomes, and explain the rationale behind the recommendations that drive tailored instruction for their students.

Activity: XAI - Mia Case Study



- ❖ **Read (5 minutes):**
 - Open the Mia AI Detection document and read it.
- ❖ **Reflect (10 minutes):** After reading, take 10 minutes to think about "explainable AI" that can clearly explain its decisions to humans in relation to this case.
 - **Transparency:** How might knowing *why* Mia flagged a cancer (e.g., a six-millimetre patch) help doctors trust and use its diagnosis?
 - **Patient Impact:** Why could it matter to patients like Barbara that Mia's decision is explainable, not just accurate?
 - **Challenges:** What might make it hard for Mia to explain its findings in a way doctors and patients understand?
- ❖ **Complete:** Complete the activity and upload to your MST Team.

Activity: Recommender



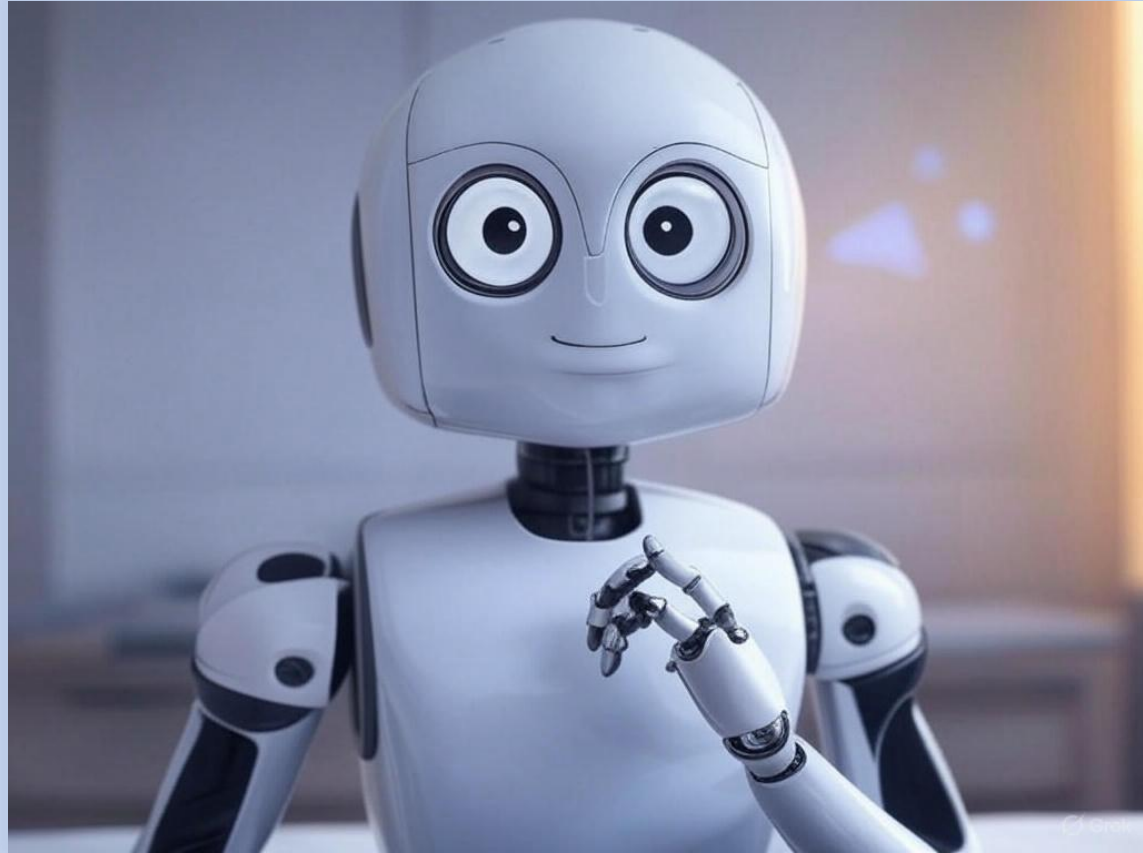
- ❖ **Ask for a Movie Recommendation:** Students use Grok's "think" feature to request a movie suggestion (e.g., "Grok, think: What movie should I watch tonight, and why?") and observe how Grok explains its reasoning.
- ❖ **Dig Deeper into the Reasoning:** Students ask Grok for a detailed explanation of a specific aspect of the recommendation (e.g., "Grok, think: Why would the coming-of-age theme in *Lady Bird* appeal to me?") to explore its thought process further.
- ❖ **Reflect on Glass Box AI:** Students consider how Grok's clear and insightful explanations demonstrate the principles of "glass box AI" and why this transparency is important for building trust and understanding in AI systems.
- ❖ **Complete:** Complete the activity and upload to your MST Team.



Exploring DeepSearch

Grok

Hints on Using Grok



- ❖ **Features:** You can turn off "Think" mode or "DeepSearch" whenever you wish. It is not always advantageous to have these modes turned on. For example, if you want Grok to draw images, rather than describing the image it is imagining, turn **off** "Think".
- ❖ **Attachments:** You can attach many different types of documents with the paperclip.
- ❖ **Logo:** Clicking on the Grok logo returns you to the home page and a new chat. Here you will see some default prompts you can modify like Code, Analyze and Research.
- ❖ **Privacy (Ghost) Mode:** You can turn on Temporary chat or search your chat history at the top right.

Activity: DeepSearch



Ai

- ❖ **Research Challenge:** Use Grok's **DeepSearch** to find and summarize recent articles on a specific topic (e.g., Redrawing of Singaporean Electoral Boundaries in 2025)
- ❖ **Fact-Checking Debate:** Employ **DeepSearch** to gather data and opinions from the web and X, using the information to support or refute a given statement. Select any statement you wish, like "Nuclear power is not safe for Singapore"





Lesson 07

Summary of Material Covered

Lesson 07 Review

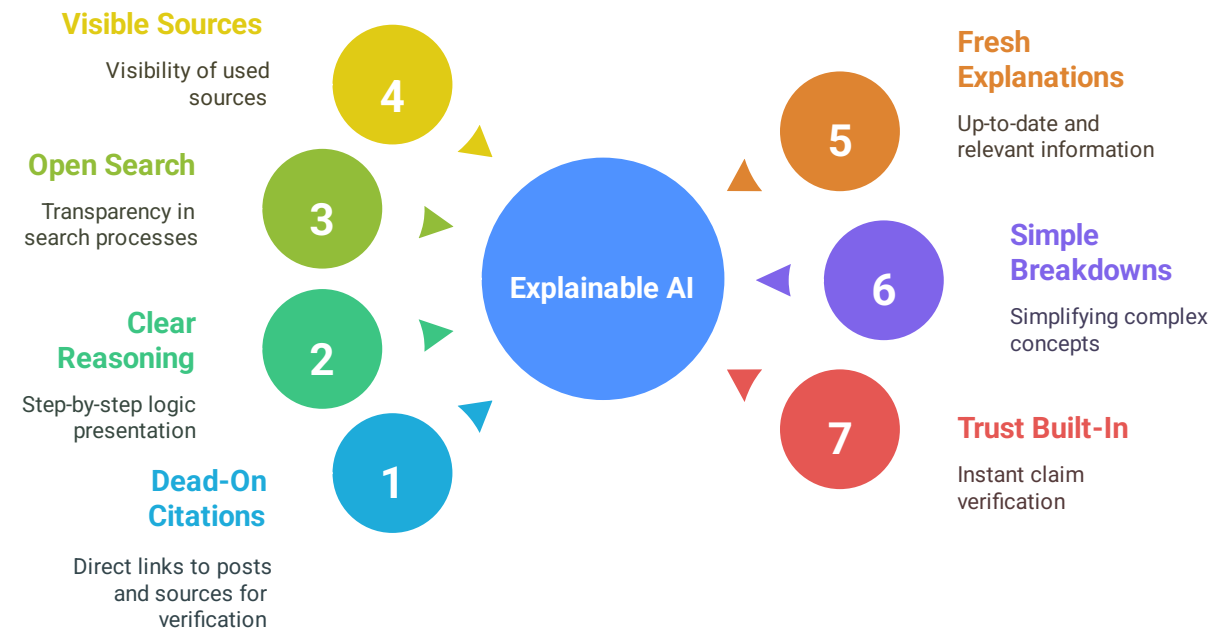
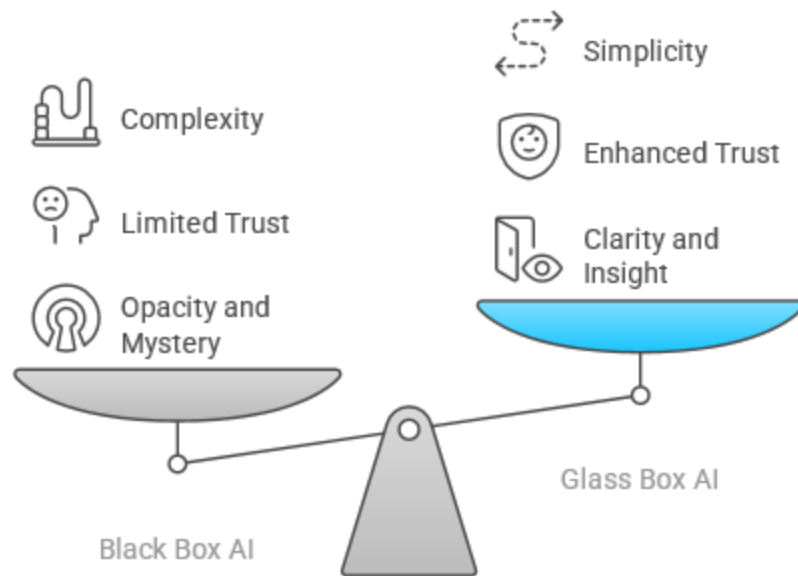


- ❖ **Intro to XAI:** Transforms AI from "**Black Box**" to "**Glass Box**" for transparency and trust.
- ❖ **Why It Matters:** Highlights trust, governance, & fairness in AI.
- ❖ **Real-Life Impact:** XAI in Healthcare, Finance, Law, & Education. Example: Mia AI in cancer detection)
- ❖ **Meet Grok:** A witty AI by X, free & interactive for exploring XAI
- ❖ **Grok's XAI Features:** How Grok supports Explainable AI. Note these features are **not** unique to Grok.
 - **Dead-On Citations:** Links straight to X posts and web sources.
 - **Clear Reasoning:** Shows its logic, step by step.
 - **Open Search:** Reveals how it digs through X and the web.
 - **Visible Sources:** See exactly what Grok uses.
 - **Fresh Explanations:** Always up-to-date and relevant.
 - **Simple Breakdowns:** Makes responses easy to grasp.
 - **Trust Built-In:** Verify claims on the spot.

Lesson 07 Review



Ai



Thank you

School of Infocomm

C240 AI Essentials and Innovations

© Republic Polytechnic 2025: All Rights Reserved

