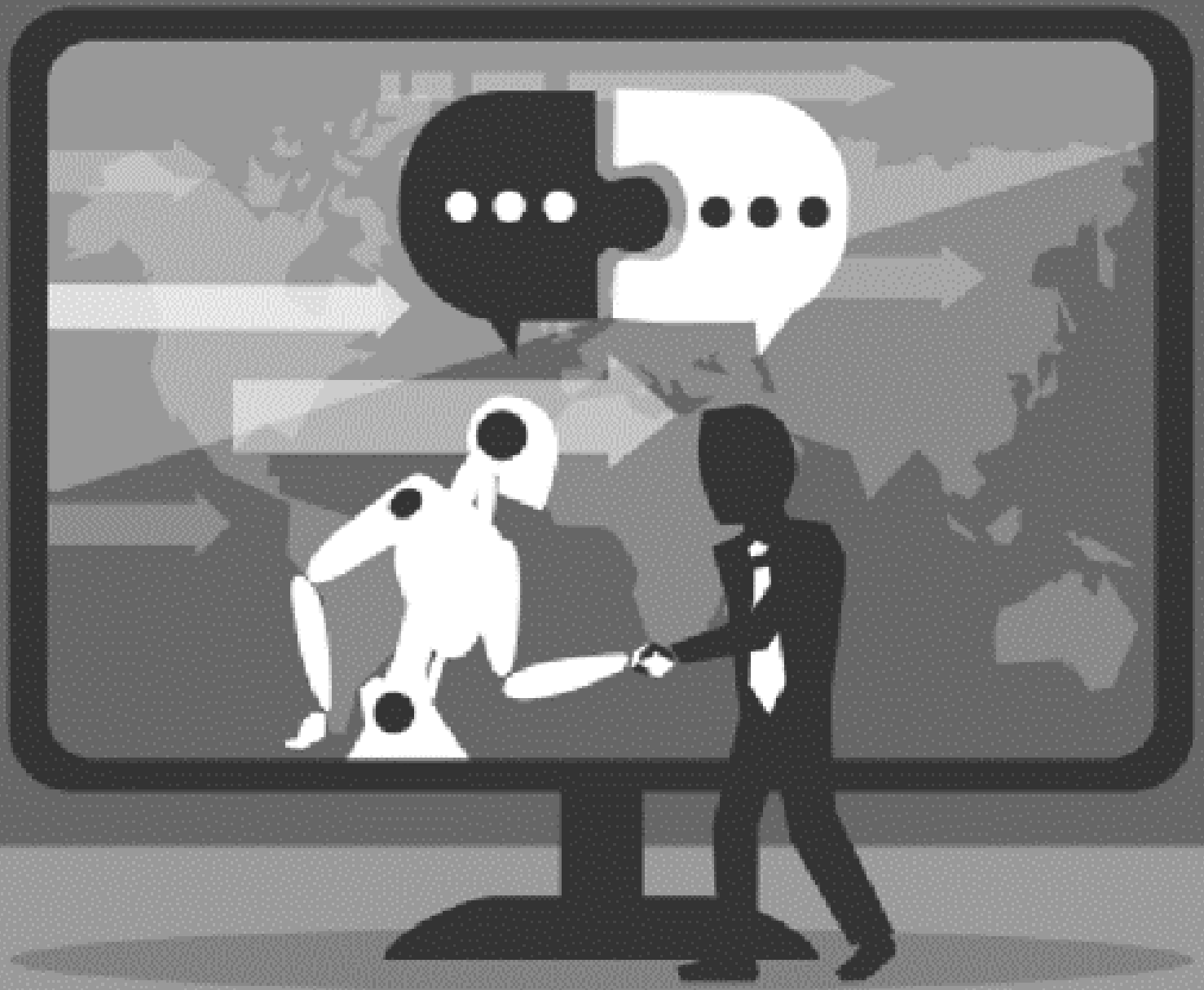
It’s no longer about what AI is — it’s about what we ask it.



Link to this folder:

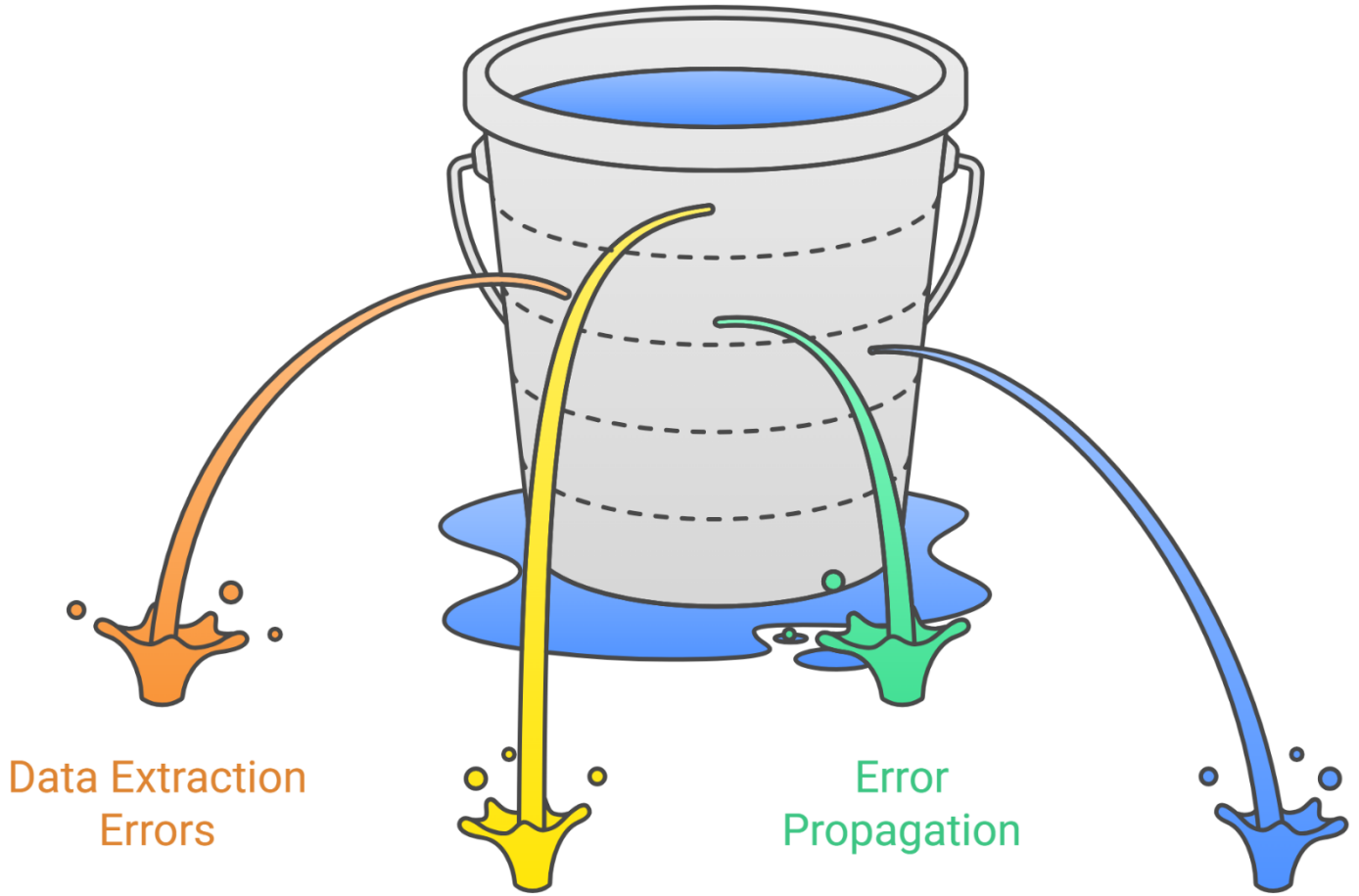
https://bit.ly/vasaiAI

https://drive.google.com/drive/folders/1OVS-mSaZS-jYe-0jl01HfQiouKkn5zz1?usp=sharing

Types of AI Hallucinations

Type of Hallucination	Definition	Comment / Risk
Factual Hallucination	When the AI states something factually incorrect or non-existent.	High risk — produces misleading or false information.
Extrinsic Hallucination	Adds information not found in the source or reference material.	Common in RAG systems when retrieval pulls unrelated or unsupported content.
Intrinsic Hallucination	Misrepresents or contradicts the actual input or source text.	Dangerous since the source exists but is distorted, leading to misinterpretation.
Logical/ Reasoning Hallucination	Faulty or inconsistent reasoning even when the facts cited are correct.	Subtle but serious in domains like tax or law where inference accuracy is vital.
Linguistic/ Semantic Hallucination	Slight misstatements or exaggerations due to poor phrasing or paraphrasing.	Alters nuance and can change the intended legal or professional meaning.

AI Hallucinations in Financial Analysis



Data Extraction Errors

Pulling inaccurate information from tables

Reliance on External Data

AI uses internet instead of provided data

Error Propagation

Errors spread through automated systems

Model Accuracy

Open-source models struggle with accuracy

Mitigation AI Hallucinations

Retrieval-Augmented Generation

AI retrieves relevant information from trusted sources

Confidence Scoring

AI assesses and indicates confidence in answers

Clarifying Questions

AI asks questions to clarify user intent

Prompt Engineering

AI structures inputs to guide accurate responses

Post-Generation Validation

AI verifies accuracy of generated responses

Continuous Feedback

AI learns from user feedback to improve

ChatGPT Agent

Traditional ChatGPT

- Conversational only
- Provides advice and answers
- Cannot interact with websites
- No file system access
- Single-step responses
- Manual implementation required

Example:
"Here's how to analyze competitors..."
(You do the work)

VS

Agent Mode

- Autonomous task execution
- Completes entire workflows
- Browses web & fills forms
- Runs code & manages files
- Multi-step problem solving
- Delivers finished results

Example:
"I'll analyze competitors for you..."
rt

Factor	Deep Research	Agent Mode
Goal	Understanding, comparing, analysing, synthesizing knowledge	Doing, executing, automating workflows
Output	Reports, summaries, analyses, insights	Actions, completed tasks, updated systems
Input	Open-ended query (“Explain impact of GST rate change”)	Specific instructions (“Login → download Form 26AS → save in Sheet”)
Risk Level	Low (only reading & summarizing)	Medium/High (can affect live data, payments, filings)
Best For	Strategy, learning, drafting, advisory	Compliance tasks, admin work, repetitive online workflows

Virtual Desktop Feature: ChatGPT's Secure Sandbox

When enabled, ChatGPT opens a secure, sandboxed desktop in the cloud where it can:

- Navigate real websites
- Download and handle files
- Fill forms and execute tasks like a human

Safety Architecture

Sandboxed & Ephemeral

Isolated environment destroyed after session ends

Permissioned Steps

Agent asks approval before consequential actions

Traceability

Full action log maintained; can stop at any step

Privacy Concerns

Credentials & Sensitive Data

Inputs go through OpenAI's servers; enter sensitive data in desktop session, not chat

Screen Visibility

Agent's actions visible in your session; potential residual exposure risk

Downloaded Data

Files temporarily stored in sandbox; deleted after session

Third-Party Sites

Risk of account lock if site detects unusual login patterns

Risk Level Assessment

Low Risk

Non-sensitive workflows (booking events, public data, form filling without personal identifiers)

Moderate Risk

Professional work with client data, but anonymized files or no shared credentials

High Risk

Directly sharing passwords, PAN, Aadhaar, bank logins, or confidential documents

Practical Precautions for CAs / Finance Professionals

Use one-time logins

✅ Use one-time logins (OTP-based, session passwords) rather than your permanent credentials.

Protect sensitive information

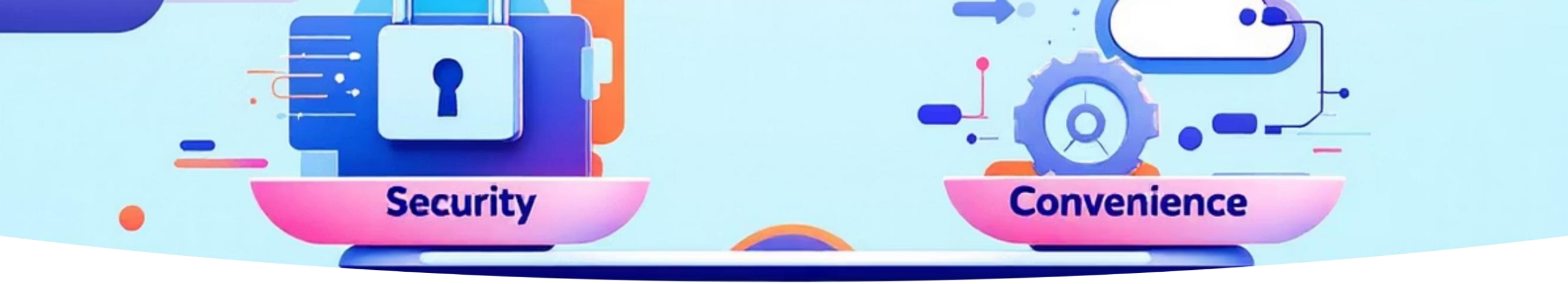
✅ Avoid typing sensitive IDs/passwords in chat. Enter them directly when the agent prompts the desktop login field.

Secure downloaded files

✅ Download client ITRs, challans, etc. into a local secure folder (or your firm's Google Drive) immediately—don't leave them in the session.

Limit access appropriately

✅ Treat the virtual desktop as a junior staff assistant—good for automation, but don't hand it unchecked access to client portals.



Bottom Line

> The virtual desktop is relatively safe for routine, semi-sensitive automation (e.g., downloading forms, filling non-financial registrations), because it's sandboxed and ephemeral.

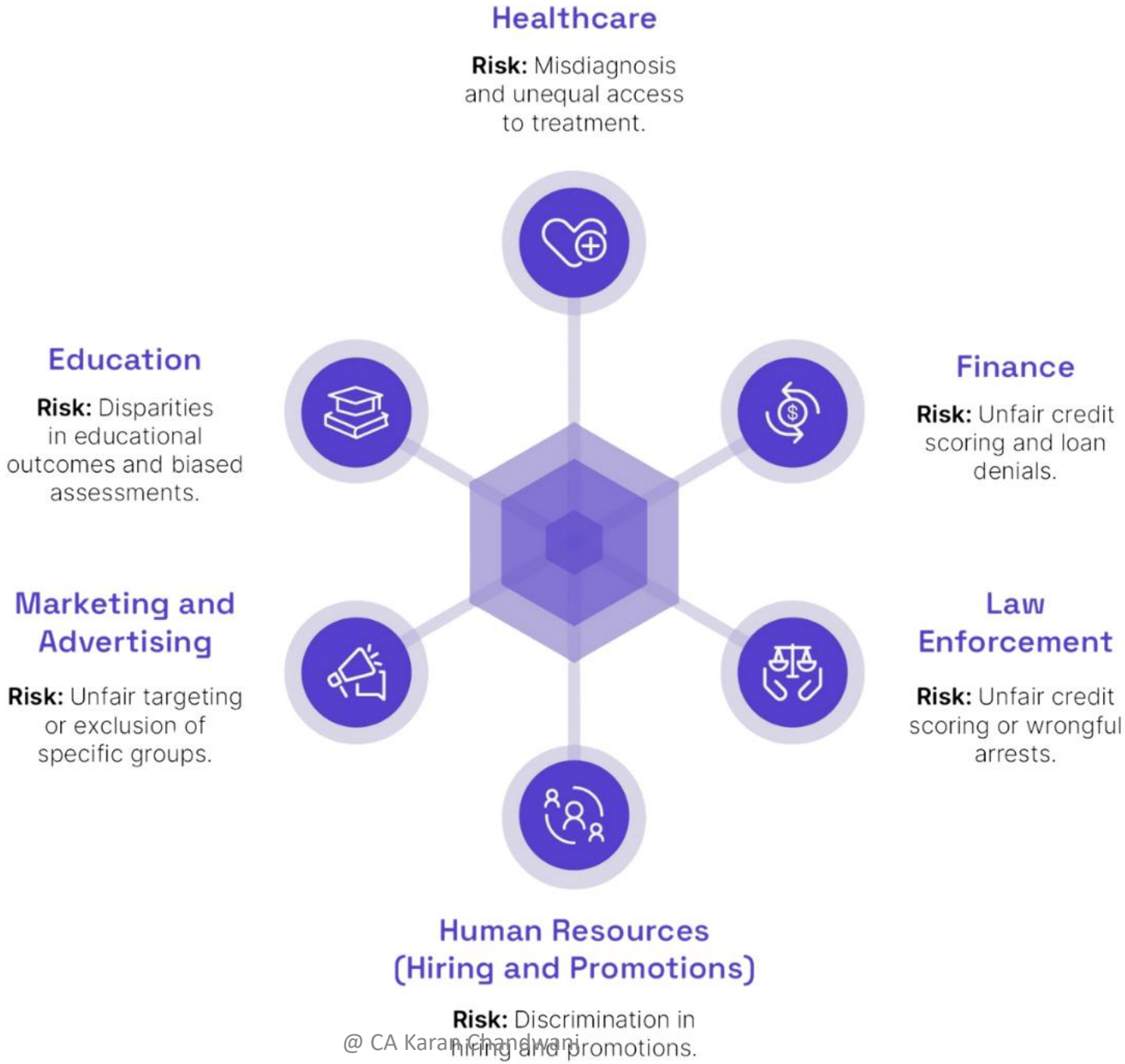
⚠ **But for critical confidential client data** (Income-tax portal, MCA filings, GST returns), you should treat it with caution—don't hand over permanent credentials, and always control the flow of sensitive information.

Remember to maintain a balance between leveraging the automation capabilities of the virtual desktop feature while implementing appropriate security measures for sensitive financial data.

🧾 Checklist for Ethical & Safe Use

✅ Do's	🚫 Don'ts
Anonymize sensitive client data	Upload personal details (PAN, Aadhar)
Use tools for draft prep only	Treat output as final/legal advice
Cross-check with trusted sources	Assume all facts are correct
Use consented data only	Upload internal documents freely
Stay updated on DPDP compliance	Ignore audit trails or logs

BIAS in AI:


Healthcare
Risk: Misdiagnosis and unequal access to treatment.
Education
Risk: Disparities in educational outcomes and biased assessments.
Finance
Risk: Unfair credit scoring and loan denials.
Marketing and Advertising
Risk: Unfair targeting or exclusion of specific groups.
Law Enforcement
Risk: Unfair credit scoring or wrongful arrests.
Human Resources (Hiring and Promotions)
Risk: Discrimination in hiring and promotions.

Case	Bias Type	Details
Amazon (2014–15)	**Gender**	Penalized resumes with “women’s” due to male-dominated training data
iTutor (2022)	**Age**	Automated rejections of older applicants (55+/60+) led to EEOC settlement
Workday (2023–)	**Age, race, disability**	Collective lawsuit over systemic rejection of older, Black, and disabled applicants
Regulation (2023–)	**ZIP code usage**	Laws ban use of ZIP as proxy to avoid discrimination against protected groups

- **Colorado AI Act (effective February 1, 2026):** Regulates the use of AI systems that make or are a substantial factor in making "consequential decisions" in areas such as employment. The law will require AI deployers (i.e., employers using AI) to use reasonable care to avoid algorithmic discrimination, implement risk management policies, complete annual impact assessments, provide notice when certain AI systems are used, and provide employees an opportunity to appeal adverse consequential decisions resulting from AI (among other requirements).

- **Illinois HB 3773 (effective January 1, 2026):** Prohibits employers from using AI in a way that results in employee discrimination on the basis of protected classes under the Illinois Human Rights Act. Requires employers to notify workers when AI is used with respect to recruitment, hiring, promotion, renewal of employment, selection for training or apprenticeship, discharge, discipline, tenure, or the terms, privileges, or conditions of employment. The law further prohibits AI systems that use zip codes as a proxy for protected classes.

- **Illinois AI Video Interview Act (went into effect January 2020):** Governs the use of AI to analyze recorded video interviews of job applicants by requiring disclosure, consent, deletion rights, and government agency reporting.

- **New York City Local Law 144 (went into effect July 2023):** Regulates the use of automated employment decision tools for hiring or promotion decisions by requiring employers to provide advance notice of such tools, conduct independent bias audits annually, and publish the results of such audits.

AI best practices moving forward

- **Implement an AI use policy.**
- **Conduct AI audits**: Regularly evaluate AI tools for potential biases against protected classes and document results.
- **Review vendor agreements**: Ensure AI vendors provide transparency regarding how their algorithms function, confirm compliance with anti-discrimination laws, and carefully scrutinize warranty, disclaimer, and indemnity provisions.
- **Implement human oversight**: AI should be used as a tool—not a sole or substantial decision-maker—for hiring, promotions, and terminations. Human review is crucial and internal policy controls should be implemented to ensure appropriate human involvement.
- **Monitor legal developments**.

AUTOMATED TAX PLANNING: WHO'S LIABLE WHEN AI GETS IT WRONG?

WHERE AI CAN GO HORRIBLY WRONG (AND COST US BIG TIME)

- **AI misinterprets tax laws (AI has one big flaw: It's not actually intelligent. It doesn't "know" anything. It just predicts what sounds right based on past information. And when it's wrong? It's still confidently wrong.) AI's tendency to oversimplify complex tax law.**

- **AI generated mistakes (misclassifies income or deductions, misapplies tax credits, pulls outdated information).**

- **AI can't handle complex tax situations (but thinks it can).**

- **Privacy and security risks (because AI stores your data).**

- **Don't forget, AI is just a tool... not a tax expert !!**

Liability When Using AI in Tax Planning

Liability Confusion

AI complicates traditional legal accountability

Legal Framework Challenges

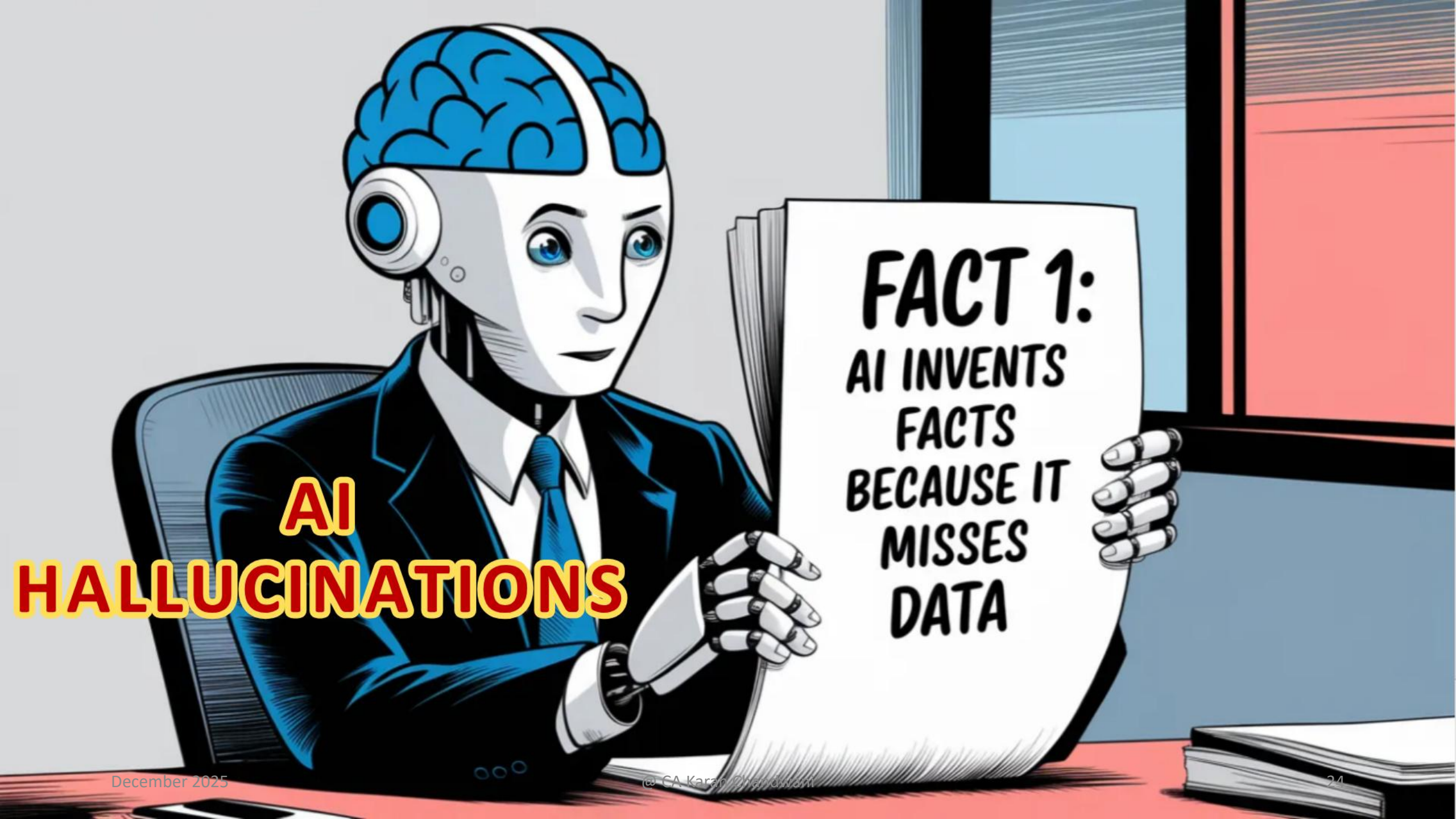
Outdated laws struggle to assign fault

Responsibility Should Lie with Humans

Professionals must verify AI tax advice

Regulatory and Ethical Recommendations

Clear liability frameworks and transparency


AI
HALLUCINATIONS
FACT 1:
AI INVENTS
FACTS
BECAUSE IT
MISSES
DATA

According to a report by Mint, the order was issued in December 2024 in the case Buckeye Trust v. PCIT-1 Bangalore (ITA No. 1051/Bang/2024) but was revoked a week later, citing "inadvertent errors."

Case details: A Rs 669 Crore trust transfer

The dispute centered around the tax implications of a transaction where an individual established a trust valued at Rs 669 crore, primarily by transferring partnership interest to the trust. Under Indian tax laws, gifts exceeding Rs 50,000 to non-relatives are generally taxable in the recipient's hands. However, the assessee's legal team argued that a partnership interest does not qualify as "property" under tax laws.

Contrary to the usual trend where such rulings favor the assessee (the Trust), the ITAT ruled in favor of the tax department. The tribunal equated partnership interest to stock market shares, deeming it taxable.

AI-generated fake judgements in the ITAT order

The ITAT's decision relied on legal precedents that were later found to be fabricated. The ruling cited the following judgments, none of which exist in legal records:

- K. Rukmani Ammal v. K. Balakrishnan (1973) 91 ITR 631 (Madras High Court)
- S. Gurunarayana v. S. Narasinhulu (2004) 7 SCC 472 (Supreme Court)
- Sudhir Gopi v. Usha Gopi (2018) 14 SCC 452 (Supreme Court) – The actual case under this citation was K. Subba Rao v. State of Telangana, unrelated to the matter.
- CIT v. Raman Chettiar (57 ITR 232 SC) – A real case concerning Hindu Undivided Family taxation, irrelevant to partnership firms.

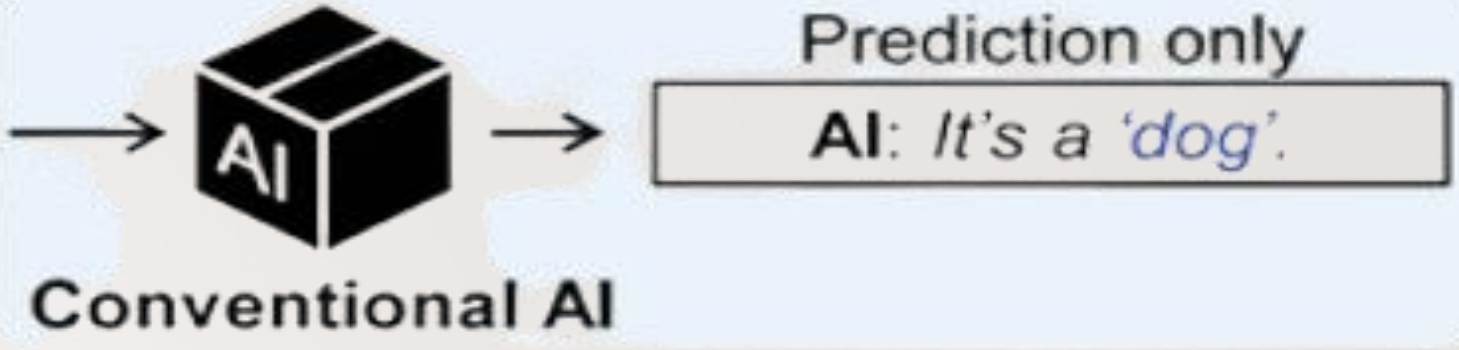
Tribunal's oversight and the role of AI

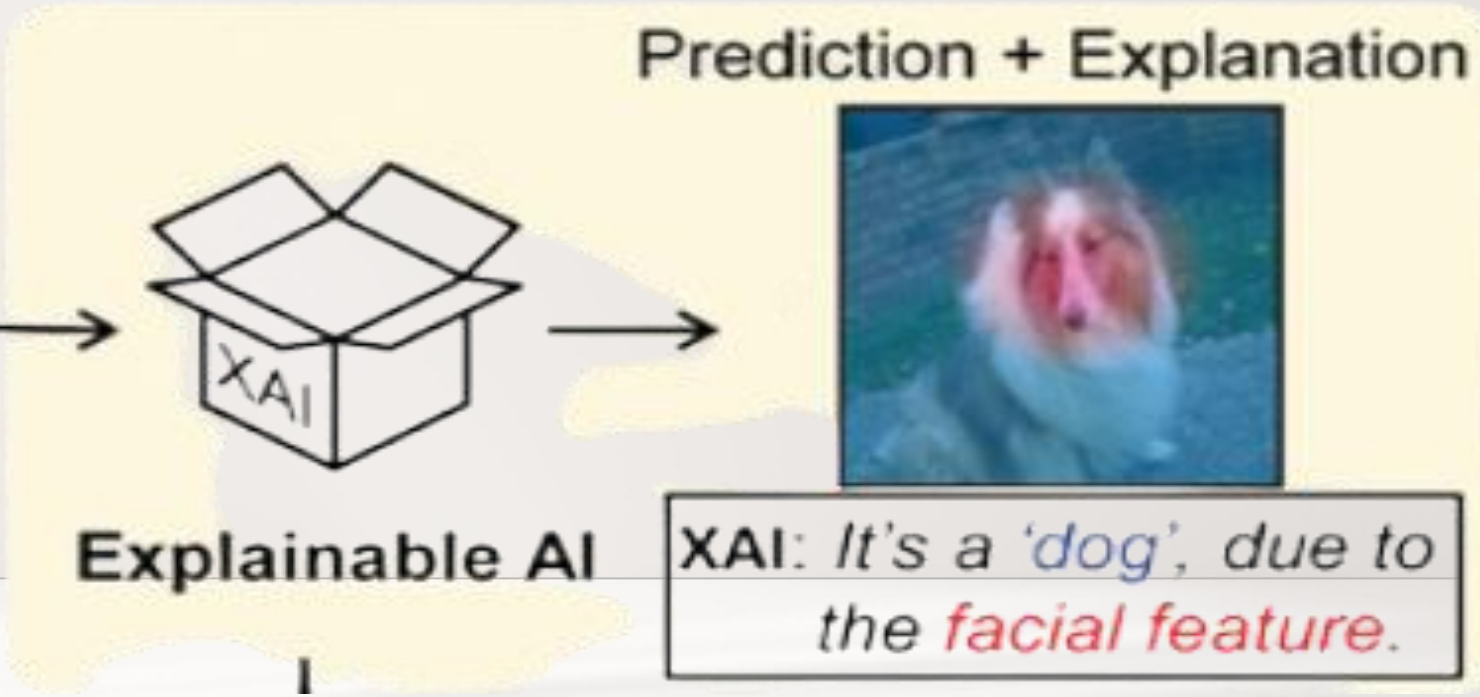
Tax department representatives reportedly used ChatGPT to generate legal precedents supporting their argument. The ITAT bench, without verifying these references, incorporated them into the ruling, leading to an erroneous judgment.

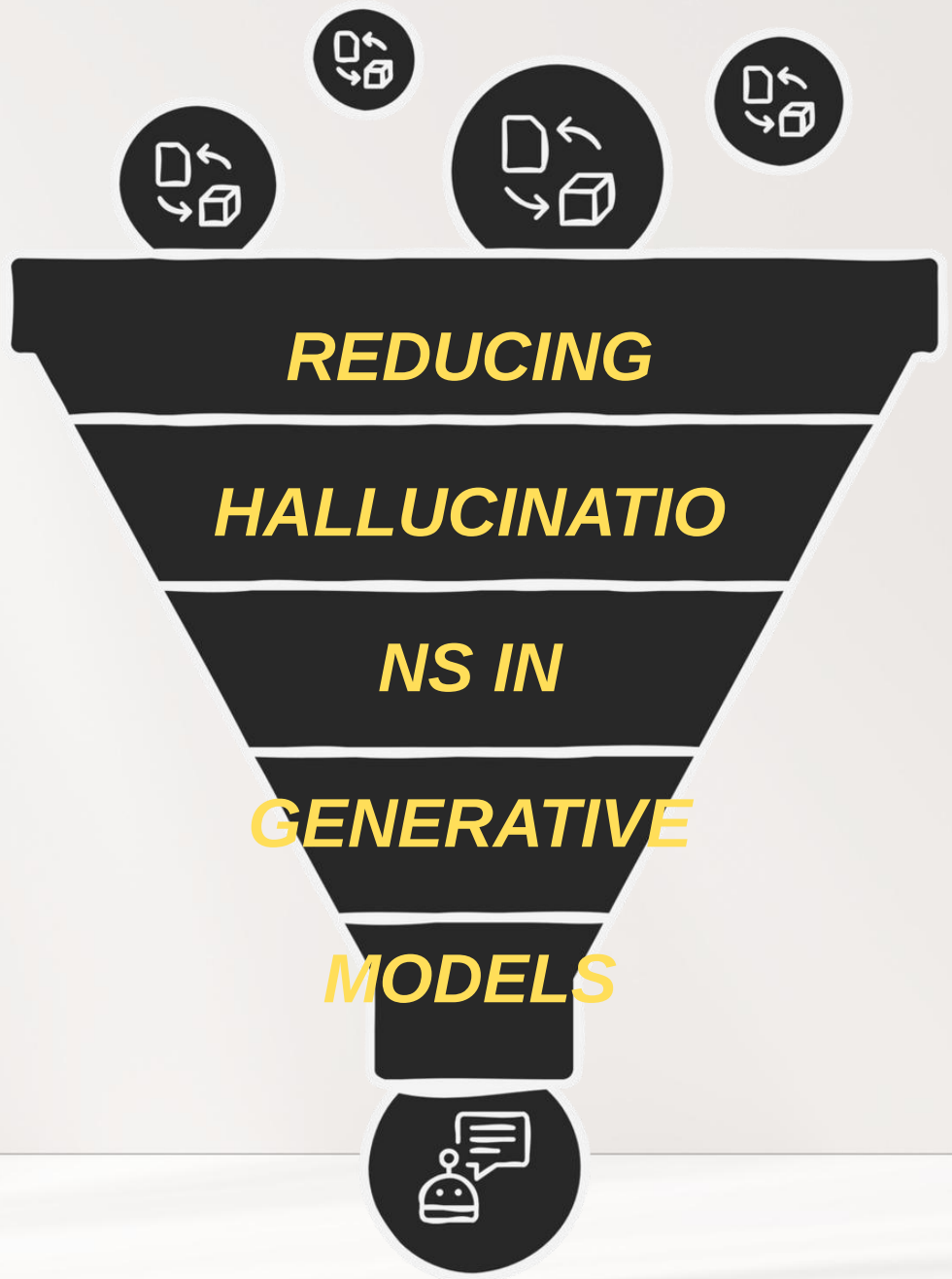
This incident highlights the growing use of AI in legal research and the risks associated with its unverified output. AI-generated "hallucinations"—false but convincing information—can lead to serious consequences if not cross-checked. The episode underscores the importance of due diligence in legal proceedings, particularly when relying on AI-generated data.

XAI – EXPLAINABLE AI

Image







Fine-Tuning

Aligning model responses with domain-specific data

Knowledge Injection

Integrating external knowledge sources for accuracy

Reinforcement Learning

Refining outputs based on human feedback

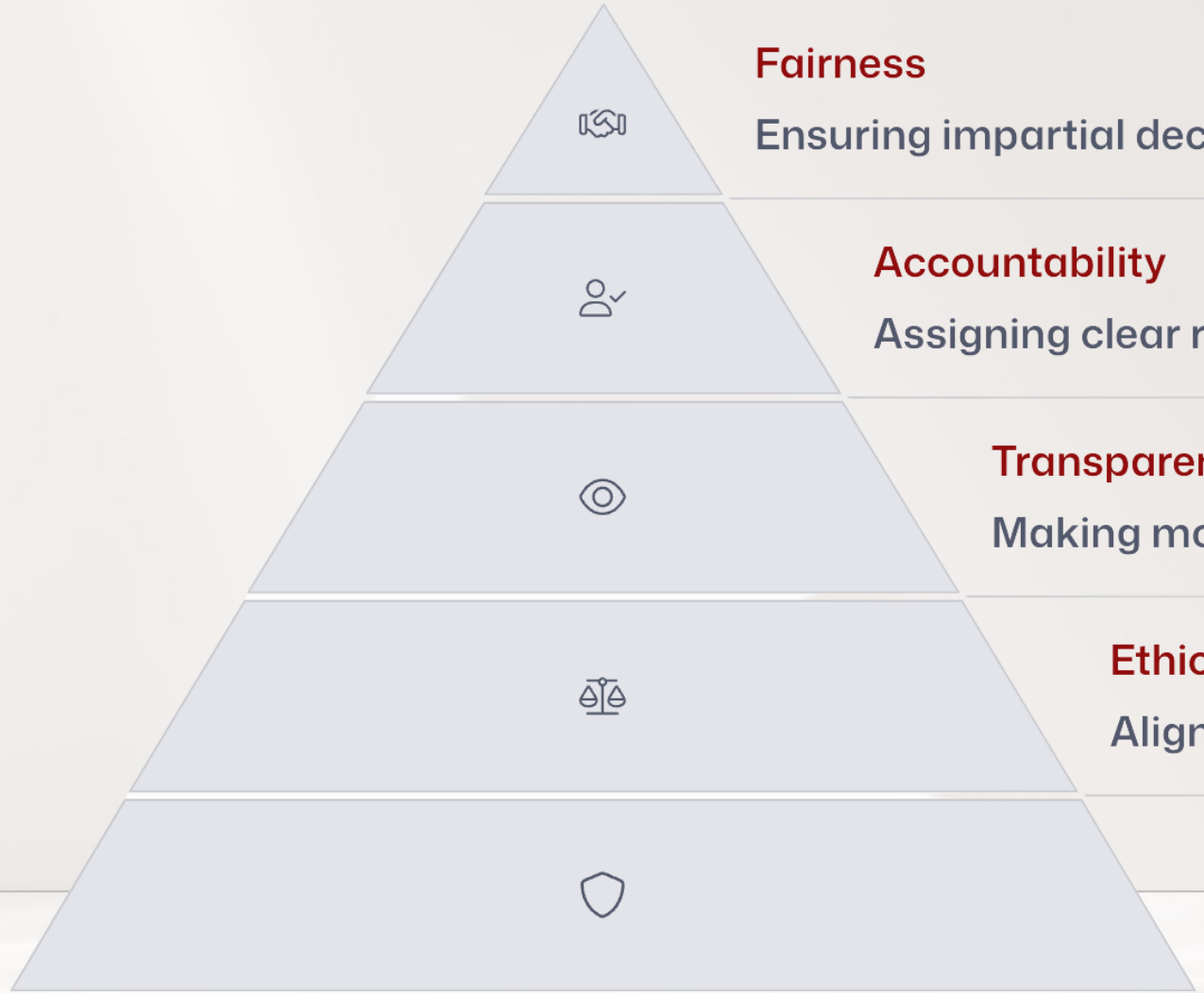
Model Ensembling

Cross-validating outputs with multiple models

Fact-Checking

Verifying and filtering outputs before release

FATE+P Framework



Fairness

Ensuring impartial decisions without unjust bias

Accountability

Assigning clear responsibility for AI decisions

Transparency

Making models and data sources open to scrutiny

Ethics

Aligning with moral principles and legal norms

Privacy

Safeguarding personal data from unauthorized access

Safeguards

Fairness Safeguards

Client segmentation audit & bias remediation process

Accountability Measures

Designated AI owner & audit trail and versioning

Transparency Protocols

Disclosure to clients & maintain data source summary

Privacy Protection

Client consent & data anonymization

Data Security

Confidentiality protocols

The machine does not control us. We control the machine. And it is our duty to use it wisely.

Thank you!

CA Karan Chandwani

FCA, B.Com

Karan Chandwani & Co.

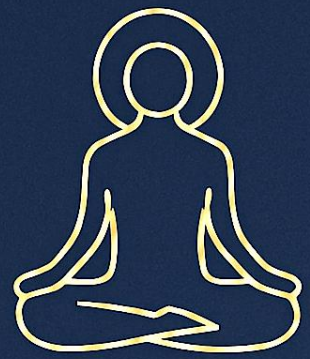
Chartered Accountants

9, Grace, Jay Kay CHS,

322/17 Shankar Shet Road, Pune 411042.

+91 97649 13651

kdchandwani1@gmail.com



Rooted in wisdom
Rising through AI