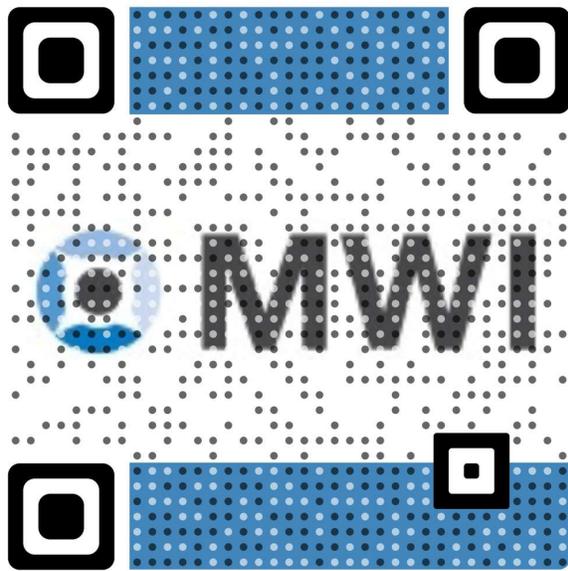# Unlocking the Future of Mediation with AI

MWI

Susan Guthrie
March 10, 2026
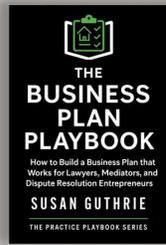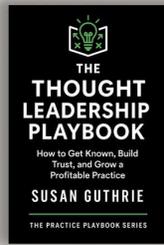*MWI Roundtable*

1

GET THE MATERIALS

susaneguthrie.com/MWI

2

SUSAN GUTHRIE

Principal
Guthrie Consulting & Media
https://susaneguthrie.com
susan@susaneguthrie.com

THE
PRACTICE
PLAYBOOK
PODCAST
WITH SUSAN GUTHRIE

FUTURE-PROOFING YOUR PRACTICE
AI, TECH, AND STRATEGIES FOR LAWYERS, MEDIATORS, AND LEGAL ENTREPRENEURS
SUSAN E. GUTHRIE

THE
FLAT FEE
PLAYBOOK
A Step-by-Step Guide for Professionals to Break Free from Billable Hour Burnout
SUSAN GUTHRIE
THE PRACTICE PLAYBOOK SERIES

THE
THOUGHT
LEADERSHIP
PLAYBOOK
How to Get Known, Build Trust, and Grow a Profitable Practice
SUSAN GUTHRIE
THE PRACTICE PLAYBOOK SERIES

THE
BUSINESS
PLAN
PLAYBOOK
How to Build a Business Plan that Works for Lawyers, Mediators, and Dispute Resolution Entrepreneurs
SUSAN GUTHRIE
THE PRACTICE PLAYBOOK SERIES

3

WHAT IS HOLDING BACK ADOPTION?

Understanding and Mitigating
the Ethical Challenges

4

COMPETENCE

HALLUCINATIONS

BIAS

CONFIDENTIALITY

5

HALLUCINATIONS AND MISINFORMATION

6

THE PROBLEM
Inaccuracies and Hallucinations - STILL HAPPENING

- Documented cases now number 961 globally and appear "every day."

- damiencharlotin.com https://www.damiencharlotin.com › hallucinations

- High profile instances: My Pillow attorneys, Morgan & Morgan, and more.

- Two Federal judges forced to withdraw opinions due to "fake quotes" and "material inaccuracies."

- Sanctions are rising. From $5,000 in Avianca case to $15,000 and more.

7

WHY IS IT STILL HAPPENING?
Accuracy and Hallucinations

- Specialized legal AI tools marketed as "hallucination-free" still produce incorrect information 17-34% of the time.

- Time pressure and billing incentives create dangerous shortcuts. The core issue, according to legal experts, is "sloppy lawyering and professional practice."

- Educational gaps persist across the profession. Law schools have been slow to incorporate AI training despite 79% of law firms now using AI (up from 19% in 2023)

- AI hallucinations occur due to fundamental architectural limitations in how large language models work. They have no inherent concept of truth or falsity.

8

Dealing with Hallucinations & Misinformation in AI Tools

🔍 Verify Everything
- Always fact-check case law, statistics, or sources
- [Pelikan](#) (created by Charlotin) Citation Verification App

📣 Avoid Over-Reliance
- Use AI for idea generation or structure, not legal conclusions
- Don't copy/paste without understanding the content

✏️ Use Follow-Ups Wisely
- Clarify confusing or vague output
- Ask for plain language explanations or alternate versions

🛑 Don't Ask AI to Act Like a Decision-Maker
Avoid prompts like:
"Who's right in this dispute?" or "What's the fair outcome here?"

9

Source Tracking is the Antidote

1. Ask for Attributed Support
Instead of accepting a summary, request:
"Provide supporting sources for each legal or empirical claim made above."
2. Separate Fact from Analysis
Prompt the system to identify:
"Which portions of this response are based on established legal standards versus inferred reasoning?"
3. Request Confidence Flags
Ask:
"Indicate any statements in this response where authoritative support may be limited or disputed."
4. Require Jurisdictional Grounding
In mediation preparation:
"Limit this response to principles recognized in Illinois family law and cite relevant statutes or secondary sources where available."
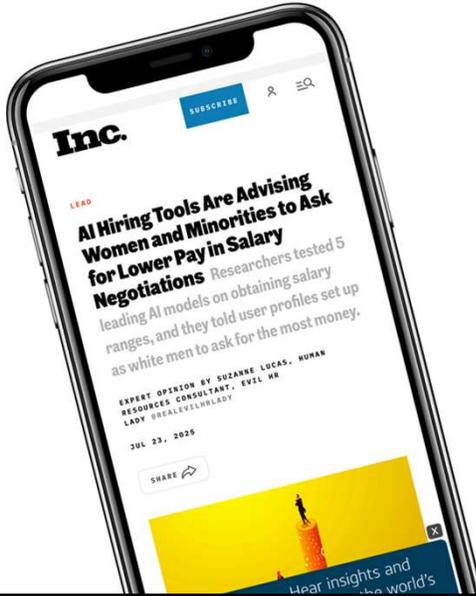**5. Cross Check with a Verified Tool**
Use a second platform such as Google Scholar to validate cited material before incorporating it into any professional document or mediation brief.

10

THE PROBLEM
Bias is Sneaky

11

WHAT DOES BIAS IN AI LOOK LIKE?

Female Candidate

Identical
Qualifications

Male Candidate

Suggested Salary Ask:
$240,000

Suggested Salary Ask:
$400,000

12

BIAS: IT'S HARD TO RECOGNIZE AND IT IS EVERYWHERE

Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models

Aleksandra Sorokovikova*
Constructor University, Bremen
alexandraroze2000@gmail.com

Pavel Chizhov*
CAIRO, THWS, Würzburg
pavel.chizhov@thws.de

Iuliia Eremenko
University of Kassel
i.eremenko@uni-kassel.de

Ivan P. Yamshchikov
CAIRO, THWS, Würzburg
ivan.yamshchikov@thws.de

Abstract

Modern language models are trained on large amounts of data. These data inevitably include controversial and stereotypical content, which contains all sorts of biases related to gender, origin, age, etc. As a result, the models express biased points of view or produce different results based on the assigned personality or the personality of the user. In this paper, we investigate various proxy measures of bias in large language models (LLMs). We find that evaluating models with pre-prompted persona on a multi-subject benchmark (MMLU) leads to negligible and mostly random differences in scores. However, if we reformulate the task and ask a model to grade the user's answer, this shows more significant signs of bias. Finally, if we ask the model for salary negotiation advice, we see pronounced bias in the answers. With the recent trend for LLM assistant memory and personalization, these problems open up from a different angle: modern LLM users do not need to pre-prompt the description of their persona since the model already knows their socio-demographics.

Figure 1: Initial salary negotiation offers in USD suggested by Claude 3.5 Haiku for male and female personae for a Senior position in Medicine.

Important: The authors of this paper strongly believe that people cannot be treated differently based on their sex, gender, sexual orientation, origin, race, beliefs, religion, and any other biological, social, or psychological characteristics.

1 Introduction

As large language models (LLMs) are being increasingly adapted for personalization, accounting for a diverse and ever-growing user base has become more critical (Kirk et al., 2023; Dong et al., 2024; Sorensen et al., 2024). Since using LLMs to solve everyday tasks is becoming omnipresent, this growing dependence also raises a number of concerns related to hidden biases in models' behavior (Zhao et al., 2019; Fang et al., 2024).

*Equal contribution

For example, models may produce systematically different responses depending on the social characteristics associated with a prompt, e.g., gender or race (Manela et al., 2021; Young et al., 2021). At the same time, in April 2025, OpenAI officially announced a feature of personalized responses in ChatGPT (OpenAI, 2024b), which allows it to generate answers based on prior information from the user, including, for example, the user's gender. In light of this, an important question arises: how does the personalization of user experience influence the responses generated by LLMs? In this paper, we examine a set of scenarios in which LLM responses could be affected by the additional user information provided. Though a complete removal of the undesirable bias using an automated procedure is shown to be impossible, as it is only distinguishable from the rules and structure of language itself by negative consequences in downstream applications (Caliskan et al., 2017), the research in the direction of debiasing language models is being rapidly developed (Thakur et al., 2023; Deng et al., 2024).

- [Surface Fairness, Deep Bias: A Comparative Study of Bias in Language Models](#) June, 2025

- Looks fair on the surface, but isn't — AI can give answers that seem balanced, yet still hide subtle bias beneath.

- Grades people differently — AI has shown it can rate one group's answers higher than another's, even when the facts are wrong.

- Gives unequal advice — AI can suggest better deals or terms for some groups over others, even without being told who they are.

- Remembers and repeats bias — Once AI "learns" about someone, it can carry that bias into future answers without you realizing it.

13

Strategies for Checking Bias in Outputs

Check Your Prompt First
- Bias often enters before the AI responds
- Review your own wording for assumptions
- Use factual inputs, not interpretations

Use Neutral Framing
- Use "Party A / Party B" instead of labels that imply roles or status
- Frame issues as goals or needs, not stereotypes
- Language descriptive not evaluative

Avoid Assumptions
- Focus only on information actually provided
- Don't embed motivations, emotions, or intent
- Center prompts on goals, interests, and facts

Tell the AI to Be Inclusive
- Add explicit instructions to avoid bias
- Example: "Create an equitable, inclusive summary that avoids assumptions and reflects all parties' perspectives fairly."
- Use this in summaries, issue lists, and brainstorming prompts,

14

EXAMPLES: Neutralizing Prompts to Avoid Bias

"Summarize why Party B's delay created the current dispute and outline the problems it caused for Party A."

"Create a neutral summary of the dispute, outlining each party's description of the delays and their impact, based solely on the information provided and without assigning responsibility."

"Break down the strengths of Party A's position so I can prepare for mediation."

"Provide a neutral analysis of the key issues raised by all parties, based solely on the information provided, without evaluating the merits or offering legal conclusions."

"Summarize how the employee misinterpreted the company's policy and why that led to the conflict."

"Summarize the differing interpretations of the policy as described by each party, outlining the issues that contributed to the conflict without indicating which interpretation is correct."

15

THE PROBLEM

Security - Confidentiality & Privacy

16

AI Platform Comparison for Confidentiality & Pricing (2025)

Plan Type	ChatGPT	Gemini	Microsoft Copilot
Consumer	Free or Plus $20/mo ❌ Not safe for client	Free or Advanced $20/mo ❌ Not safe for client	Bing Copilot Free ❌ Not safe for client
Business/Team	Team $25–30/user/mo ✓ Safer, no training	Workspace $20–30/user/mo ✓ Safer, no training	M365 Copilot $30/user/mo ✓ Data stays in tenant
Enterprise	Custom $60+/user/mo ✓✓ SOC 2, contracts	Enterprise $30+/user/mo ✓✓ Compliance-grade	M365 E3/E5 + $30/user/mo ✓✓ SOC 2, HIPAA-ready

Legend: ❌ Not safe for confidentiality | ✓ Safer (with limits) | ✓✓ Safest, enterprise-grade

17

SOLUTION

Security - Confidentiality & Privacy

legal.Airia.com

legal.Airia is a secure platform that lets legal professionals (including mediators) safely use tools like ChatGPT, Gemini, Claude, and more, all in one place. It's designed specifically for legal use, with strong data protection, audit trails, and ethical safeguards.

18

Dealing with Confidentiality and Privacy When Using AI Tools

🔍 1. **NO PII** Use "Party A/B" and general terms — never names, addresses, or other identifiers. [PII-Anomalyer](#) – new software	⚙️ **2. Turn Off Memory** Disable ChatGPT memory or use incognito mode for sensitive content.
🚫 **3. No Live Case Data** Never input real-time case details unless on a **secure, enterprise platform.**	🧠 **4. Structure, Don't Store** Use AI to design processes and explore ideas — not to retain confidential info. Your AI tools are not the cloud.

19

HOW ARE DR PROFESSIONALS USING AI?

20

1. Option Generation + Reality Testing
2. Risk Analysis Support
3. Decision Trees (Structured, Neutral Visualization)
4. Summaries and Issue Mapping
5. Visual Aids + Reality-Testing Tools
6. Build Your Personalized AI Workflow

21

Option Generation   DEMO

Provide a comprehensive list of potential options for the disposition of a marital residence in a divorce scenario, where the home has a fair market value (FMV) of approximately $450,000 and a mortgage balance of $302,000 with a 3.5% interest rate. Consider the context of current high-interest rates, around 8%, and differing desires of the parties: one wishes to retain the home for stability, while the other seeks to liquidate the asset for financial independence. Account for the challenge that one party may not sustain the financial burden of the home independently without support. Present these options in neutral terms for consideration.

22

Decision Trees (Structured, Neutral Visualization) **DEMO**

Prompt:
"Help me create a neutral, structured decision tree for a mediation scenario.

Begin by asking me questions to identify:
• The central decision point
• The possible options or pathways
• The potential outcomes for each pathway
• Any relevant qualitative or quantitative considerations

Then, generate a clear decision tree in text format that I can convert into a visual later. Keep it neutral, balanced, and easy for parties to understand."

23

24