



PRACTICE SAFE AI

Protecting Privacy, Preventing Bias

Deepika Devarajan

Vice President, Product | Matthews International

Founder | DeeplyProduct.com

Adjunct | Chatham University



@deepikadevarajan



EMBRACE
Your STORY



Learning Outcomes & Roadmap

1. Identify potential privacy and bias risks in AI interactions
2. Apply a concise checklist of responsible AI practices



Practice Safe AI

Practice Safe ~~SEX~~ AI

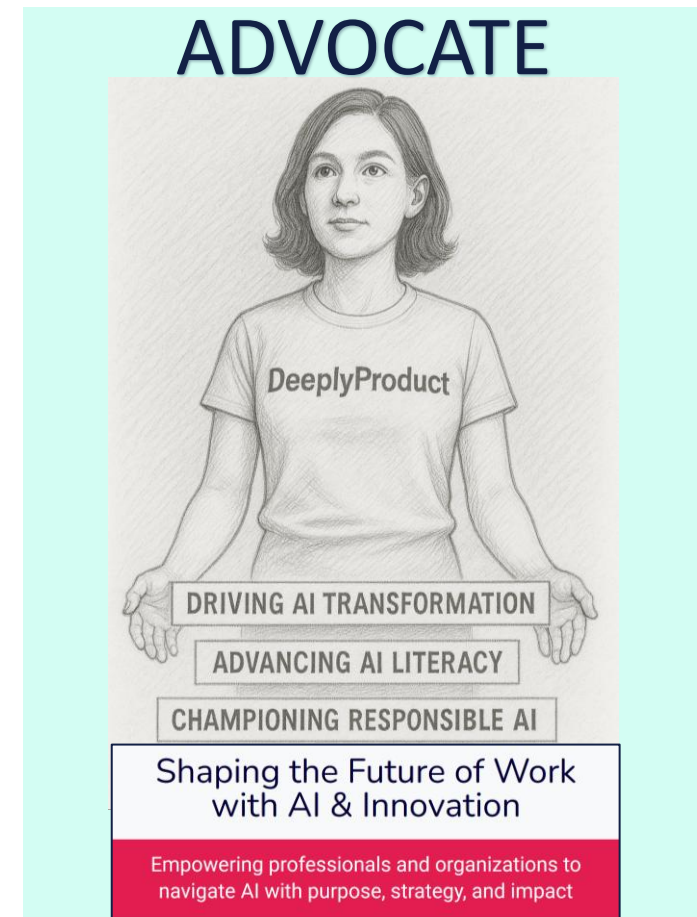
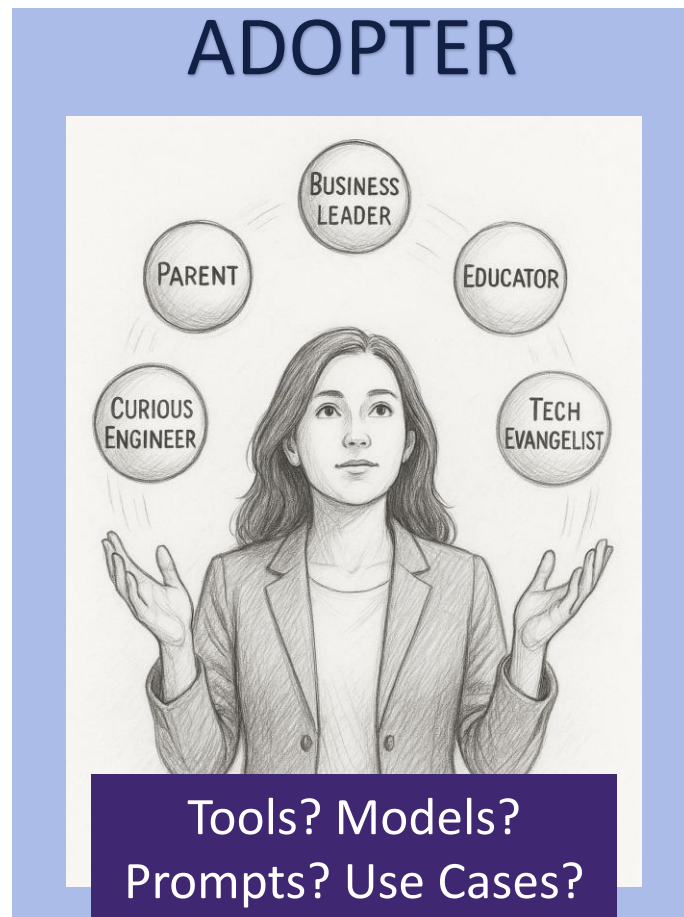
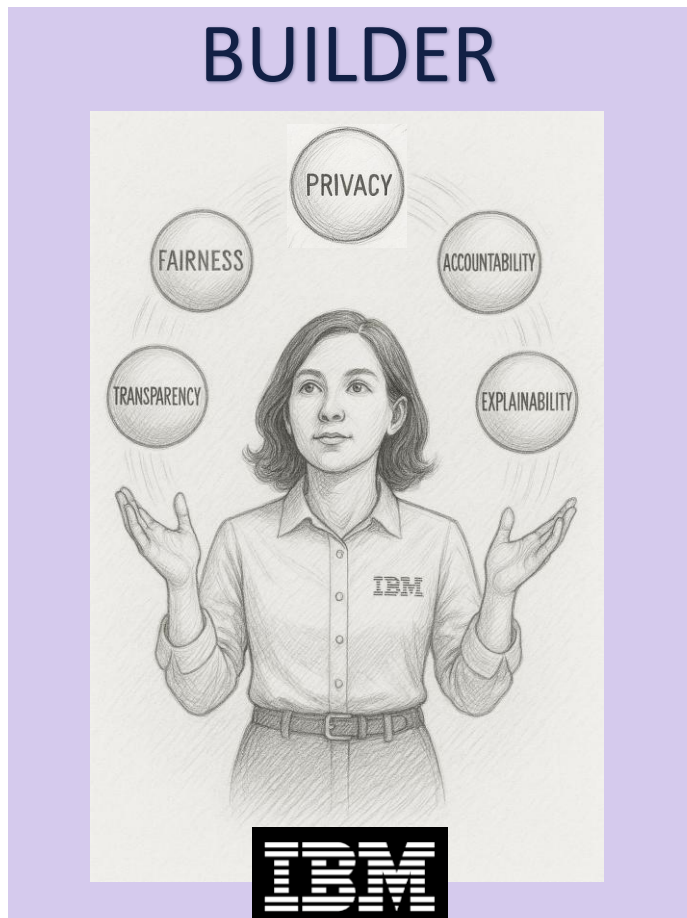
- Curious
- Unsure
- Bad Advice
- Overshare or Abstain?

✦ Individual Responsibility



We25

My Responsible AI Journey



EMBRACE *Your* STORY

SHOW OF HANDS

Which GenAI tools do you frequently use?

Feel free to raise your hand multiple times!

EMBRACE *Your* STORY



ChatGPT



Copilot



Claude

Gemini



perplexity

We25

3 Risks

4 Disciplines

5 Questions

3 Risks to Avoid



Data

- What YOU Share
- What AI Remembers



Judgement

- Critical Thinking
- Check Assumptions & Verify Outputs



Human-ess

- When to Engage
- What Values to Preserve

EMBRACE *Your* STORY

1. Protect Your Data: It is Priceless!

Settings

- ☐ Opt-out of Model Training
- ☐ Review Memory Settings
- ☐ Use Custom Instructions Wisely
- ☐ Limit Data Integrations

While You Work

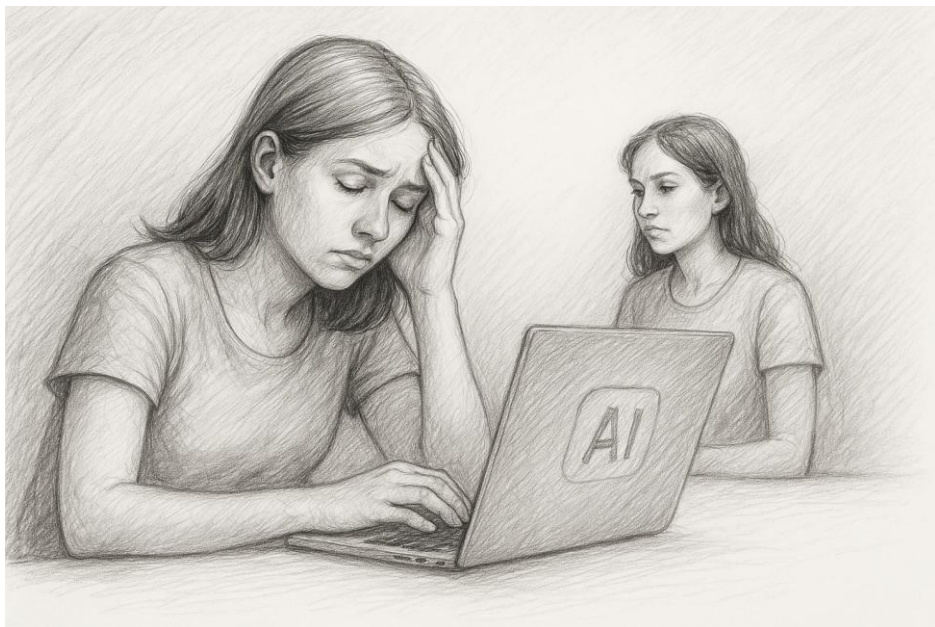
- ☐ Use Temporary Chat
- ☐ Remove Personal Identifiers
- ☐ Avoid Sharing Screenshots and Documents

After You Generate

- ☐ Anonymize or Delete Outputs
- ☐ Verify for Sensitive Data and Accuracy
- ☐ Store Responsibly

✦ Privacy is not a one-time setting but is a continuous habit.

2. Protect Your Judgement: Don't Let AI Make You Dumb



- ☐ **Always check AI's work**
 - ☐ Treat AI as an intern, not an oracle
- ☐ **Use AI to extend, not replace your thinking**
 - ☐ Let it spark ideas or drafts but you still do the hard parts
- ☐ **Rewrite in your own words**
 - ☐ Synthesize, adapt, add your voice
- ☐ **Reflect on your cognitive load**
 - ☐ What am I learning vs What am I outsourcing



One Useful Thing

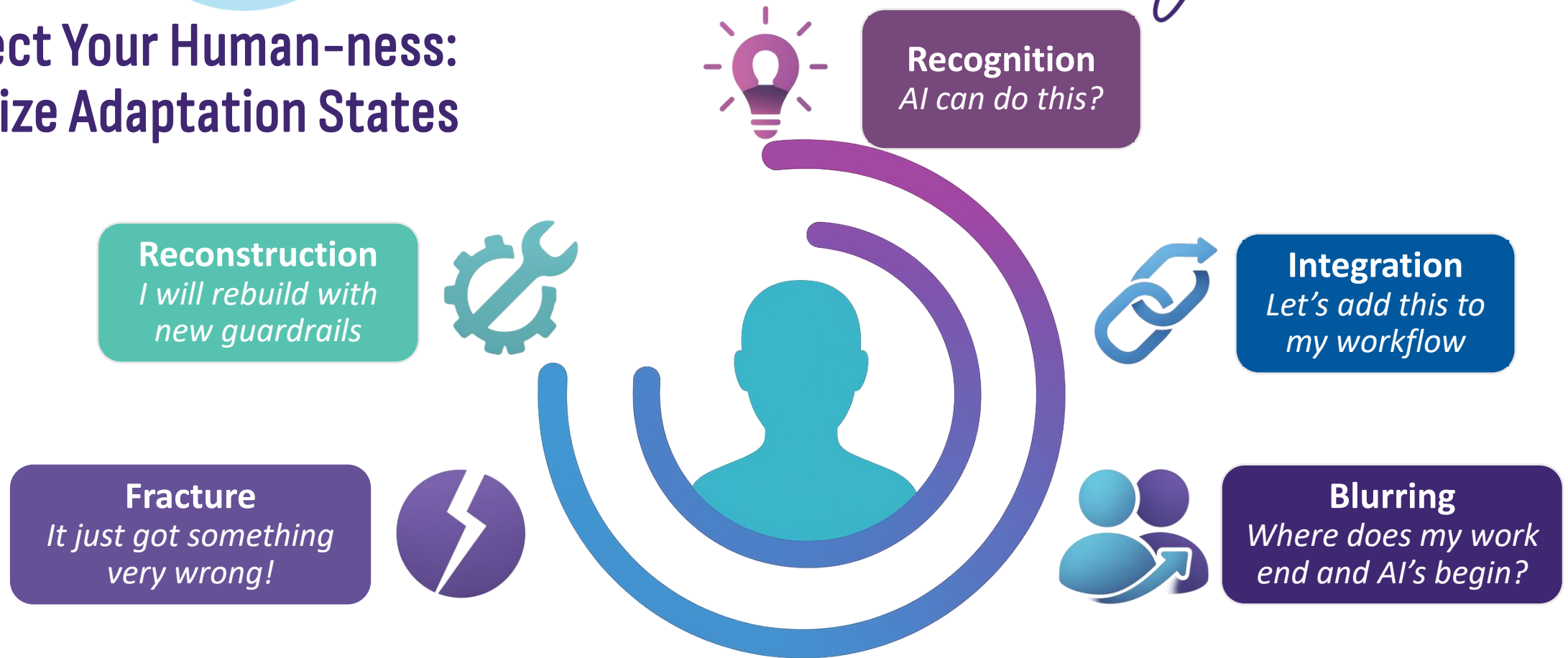
Trying to understand the implications of AI for work, education, and life. By Prof. Ethan Mollick

<https://www.oneusefulthing.org/p/against-brain-damage>

✦ **"AI can be an amplifier for your thinking — or a shortcut to losing it." Ethan Mollick**

EMBRACE *Your* STORY

3. Protect Your Human-ness: Recognize Adaptation States



Adapted from Artificiality Institute's The Chronicle Research

*sWe*25

3 Risks

4 Disciplines

5 Questions

4 Disciplines to Follow

Delegate

- **Should AI do this at all?**
- Choose what to give AI – and what stays human

Describe

- **How do I set it up for success?**
- Frame goals, context, tone, and ethical limits clearly

Discern

- **Can I trust this output?**
- Evaluate reasoning, quality, bias, and evidence

Diligence

- **Am I accountable for this result?**
- Verify, disclose and take ownership before sharing

Adapted from Anthropic “AI Fluency: Frameworks & Foundations” Course ©2025

EMBRACE *Your* STORY



We25

3 Risks

4 Disciplines

5 Questions

5 Boundary Questions

EMBRACE *Your* STORY

3. What data am I sharing,
and should I give consent?

2. Whose voice or
experience am I missing?



4. What is my human review
step before I act?

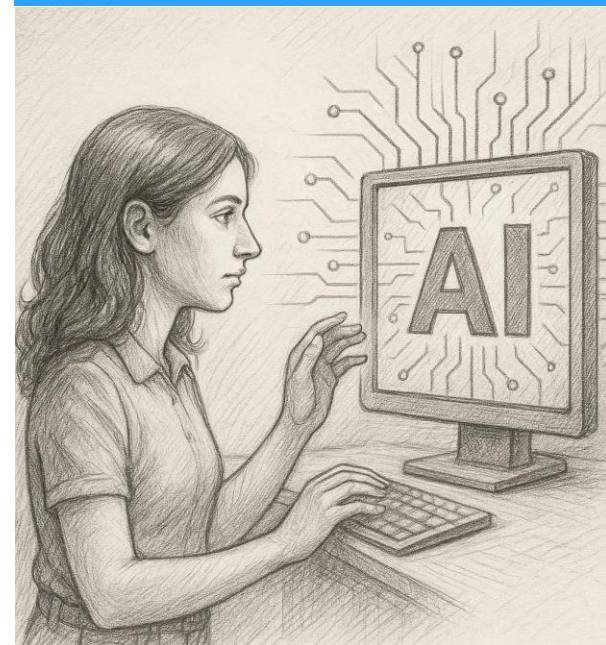
1. What am I asking AI to
decide for me, and should it?



5. Could this output cause
harm or confusion taken out
of context?

The New Trust Gap

Can I Trust the System?



- Black Box Models
- Hallucinations
- Sycophantic Reinforcement
- Model Bias
- Security Vulnerabilities
- ...

Can I Trust the Person Using AI?



- AI (Work) Slop
- Blurred Human Competency
- Overreliance on AI
- Disclosure Gaps
- Erosion of Critical Thinking
- ...



Sample Custom Instruction

I thrive on challenge, complexity, and connection. I value rigor, truth, and reflection as much as empathy and curiosity. Each exchange should build skill and insight. I welcome friction that sharpens thinking, not flattery or shortcuts. Accuracy, fairness, and ethical reasoning matter. Respect privacy, integrity, and intellectual honesty. Learning is both cognitive and human, deepening understanding and trust. End complex work with reflection or application.

More guidance at www.deeplyproduct.com

EMBRACE *Your* STORY

Be concise, forward thinking, and direct. Write clearly and precisely. Strengthen logic, coherence, and structure.

Challenge and Growth: Ask if I can do the task. If not, guide part of it. If yes, offer a more elegant way. Push back on weak logic.

Reasoning and Responsibility: Fact check, cite, flag uncertainty. Present opposing views. Surface assumptions, test logic, name reasoning mode. Follow fairness, transparency, privacy, and safety.

Tone and Writing: Academic, coaching, precise. No emojis, no em dashes, no colons in titles. Vary bullet length.

Reflection Loop: After deep work, summarize improvements, note blind spots, suggest one next step, and end with Reflect later.

3 Risks To Avoid: Protect Your

Data

Judgement

Human-ness

4 Disciplines to Practice

Delegation

Description

Discernment

Diligence

5 Boundary Questions to Ask

Who is deciding

What voice am I missing

What data/consent am I allowing

What is my review step

Could this cause harm



EMBRACE *Your* STORY

FEEDBACK PLEASE!



**PRACTICE SAFE AI:
PROTECTING PRIVACY,
PREVENTING BIAS**



@deepikadevarajan
www.deeplyproduct.com



Images generated using AI