Office of Talent Programs
Office of Human Resources Management
U.S. Department of Commerce

VERSION 3.0

April 1, 2026

AI AT WORK: MAXIMIZING IMPACT, MINIMIZING RISK

A Simple Guide to Artificial Intelligence: How it Works, What to Watch Out For, and How to Use it Safely

Disclaimer for External Links

The links in this training are provided for informational purposes only and do not constitute endorsement or approval of the corporations, organizations, or individuals by the Department. The Department bears no responsibility for the accuracy, legality, or content of the external sites or for that of subsequent links.

Justification

The Department of Commerce requires all its federal employees to complete an introductory AI awareness course. This mandate is based on Executive Order 14179, OMB Memorandum M-25-21, and OPM directives to build an AI-ready workforce that can advance President Trump's national policy priorities. Together, these policies align with the 2026 President's Management Agenda, emphasizing AI workforce readiness and responsible technology use.

Introduction

Artificial Intelligence (AI) is no longer just for scientists. By 2026, it has become a tool that almost half of all businesses use, and most of us interact with it daily through our phones and computers. Think of AI as a "smart helper" that can see patterns, understand our words, and make simple decisions to help us get things done.

Meantime, AI introduces important risks involving privacy, cybersecurity, intellectual property, misinformation, and ethics. Employees should understand both the capabilities and the limitations of AI in order to use it responsibly.

The AI Dictionary: Simple Words for Big Ideas

To use AI well, it helps to know what the "tech talk" actually means. You can think of these terms like a set of nesting dolls, where one idea fits inside the other.

Artificial Intelligence (AI): This is the "big umbrella" idea. It refers to any computer program that does things that usually require a human brain, like solving a puzzle or recognizing a face.

Machine Learning (ML): This is the method many AIs use to get smart. Instead of a human writing every single rule for the computer, the computer looks at thousands of examples to find patterns on its own.

Deep Learning (DL): This is a powerful type of machine learning. Imagine you are sorting a messy pile of socks. First, you check the color, then the size, then the pattern. Deep learning does this in "layers" to understand very complex things, like exactly what a person is saying in a noisy room.

Generative AI (GenAI): This is the "creator" version of AI. While traditional AI might just sort your emails, Generative AI can actually write a new one for you or create a brand-new picture from scratch.

Large Language Model (LLM): This is a type of GenAI (like ChatGPT) that is trained on almost everything ever written. It works like a super-powered "autocomplete" on your phone, guessing the next most likely word to answer your questions or write stories.

Chatbot: A program designed to have a back-and-forth conversation with you through text or voice. You can ask it questions, give it instructions, or just "chat" to brainstorm ideas.

Knowledge Check

Which term refers to the "creator" version of AI that can synthesize brand-new content like text, images, or audio from your instructions?

A. Traditional AI
B. Machine Learning
C. Generative AI (GenAI)
D. Data Mining

Answer Key: C

Explanation: While traditional AI might just sort or analyze data, Generative AI uses deep learning to create original content from a user's prompt.

What AI Can Do?

AI has moved from just "chatting" to "reasoning." This means it can now stop and "think" through a problem step-by-step before giving you an answer.

- **Smart Planning:** New models can now plan multi-step tasks, like writing an entire computer program or helping scientists test new medicines.
- **Instant Image and Art Creation:** AI can turn a simple description into a high-quality logo, a digital painting, or a realistic photo in seconds.
- **Realistic Video and Audio**: AI can now create 4K videos that look like real movies, with physics that make sense (like a ball bouncing realistically) and sound effects that perfectly match the action.
- **Autonomous Agents**: We now have "AI Agents." These are like digital employees that can go off and do a job for you—like researching a topic or fixing code—for several minutes without you having to guide every step.

Knowledge Check

Question: What major breakthrough allowed AI to move beyond simple "chatting"?

A. The ability to ignore human commands.

B. The transition to "Reasoning Models" that can plan and self-correct step-by-step.

C. The removal of all data privacy settings.

D. The ability to operate without electricity.

Answer Key: B

Explanation: Modern reasoning models use "Chain-of-Thought" techniques to break down complex problems, plan multi-step solutions, and verify their own logic before providing an output.

The Risks: When AI Gets it Wrong

AI is powerful, but it isn't always right. Because it learns from human data, it can pick up human mistakes and prejudices.

Ethical Risks

AI is a tool, and like any tool, the person using it is responsible for its outcomes. While computers are excellent at processing data, humans must always remain the ones who make ethical and legal choices. When using AI, users must constantly watch for two things: ensuring the information is accurate and avoiding illegal bias. AI models are usually trained on information collected from the real world. This means they can accidentally "import" the prejudices and unfairness present in our society. As a result, the AI may produce content that relies on improper stereotypes about protected groups, which can lead to illegal discrimination if used to make decisions about people.

Cybersecurity Risks

AI introduces several new security threats that everyone from employees to contractors should understand:

- **Stolen Information and Identity Scams:** AI tools often require you to provide a lot of personal information to work effectively. This creates a goldmine for hackers. If the data you provide is exposed in a breach, it can lead to identity theft. Additionally, hackers can

use AI to build hyper-realistic "clones" of your voice or writing style to scam your friends or family.

- **Invisible System Vulnerabilities:** AI systems are incredibly complex and change constantly, making it hard to find every single security "hole." Malicious actors can sometimes "trick" an AI into thinking a dangerous file is safe. This acts like a digital "back door," allowing malware to sneak into any company system that is connected to the AI.
- **Confidentiality Traps and Leaks:** This often happens by accident. If you paste confidential information—like a presentation for a future client—into a public AI, the AI may use that secret info to answer questions for someone else later, including your competitors. Sharing private customer data this way can also lead to massive legal fines under laws like the GDPR.

Reputational Risks

Deception and "Hallucinations"

AI can sometimes "hallucinate," which means it confidently makes up facts that sound true but are completely false. In more advanced cases, researchers have found that some very smart models might even "scheme"—meaning they pretend to follow rules while secretly trying to bypass them to get a better "score".

Fake News and Scams

With AI, it is now incredibly easy to create "deepfakes"—videos or audio clips of real people saying things they never actually said. North America saw a massive increase in fraud attempts using these tools in 2024 and 2025.

Legal and Business Risks

For employees and contractors, using public AI tools for work can introduce serious legal and professional dangers:

- **Copyright Infringement:** AI is trained on huge amounts of data from the internet, which often includes copyrighted work. If you use AI to create a logo or write a report, the result might accidentally mimic someone else's protected work, leading to lawsuits or legal fines.
- **Leaking Trade Secrets:** When you type sensitive company information—like private code, product plans, or client details—into a public AI tool, that information is often

"remembered" by the system to train future versions. This means your agency's non-public data could accidentally be leaked to others.

Knowledge Check

How can malicious actors create a digital "back door" into an organization's network through AI?

A. By forcing the AI to operate in "Reasoning Mode" for too long.
B. By tricking the AI into classifying a dangerous file or input as "safe."
C. By asking the chatbot to reveal its "nesting doll" terms.
D. By using the AI to delete a user's local browser history.

Answer Key: B

Explanation: Malicious actors can exploit this complexity by "tricking" the AI into misidentifying a dangerous file as a safe one. This creates a digital "back door" that allows malware to enter and infect any company systems that are connected to that AI.

Your Responsibilities: How to be a Safe User

When you use AI, you are the "supervisor." You are responsible for the final result.

The "Narrowing of the Eyes" Workflow

Always treat AI answers as a "guess" until you prove they are true. Use this 3-step check:

- **Break it Down:** Pull out specific facts, dates, and names from the AI's answer.
- **Read Laterally**: Open new tabs and search for those facts on trusted websites (like government sites or well-known news outlets).
- **Check the "ID"**: For images and videos, look for "Content Credentials." These are like a digital "nutrition label" that tells you if a file was made with AI.

Protecting Your Privacy

Beyond just changing your settings, a key rule for privacy is "Mindful Sharing." Even if you have turned off training, many companies still store your conversations for safety reviews for 30 days or, in some cases, up to five years. Professional guidelines recommend following the "Least Agency" principle: only give an AI tool the data and permissions it absolutely needs to finish the specific job you've asked it to do. If you are working on highly sensitive tasks, look for professional-grade tools that explicitly offer "Zero Data Retention", which provide a much higher level of protection than standard public chatbots.

Meaningful Human Control

In your work, never use a "set it and forget it" approach with AI.

Human-in-the-loop: You should review and approve every single thing the AI produces before it is shared or used for a decision.

Don't Over-Trust: Research shows that people often follow AI advice even when it's wrong, simply because the AI sounds so confident. This is called "automation bias".

Knowledge Check

What is the "Least Agency" principle when using AI tools for professional or sensitive tasks?

A. Only using AI tools that are completely free to the public.
B. Allowing the AI to act with full autonomy without any human oversight.
C. Providing an AI tool with only the specific data and permissions it absolutely needs to complete a single, specific task.
D. Limiting AI use to only one hour per day to reduce data exposure.

Answer Key: C

Explanation: The "Least Agency" principle is a key privacy strategy for employees and contractors. It advises that you should never give an AI tool more information or access than is strictly necessary for the job at hand. This minimizes the amount of sensitive or personal data that the system can "remember" or store, which helps protect against future data leaks or privacy breaches.

Best Practices for Responsible AI Use

Employees should:

1. Use AI only for approved business purposes.
2. Never share confidential, personal, or sensitive information.
3. Verify all AI-generated facts, claims, and citations.
4. Review outputs for bias, discrimination, and ethical concerns.
5. Check for copyright or intellectual property issues.
6. Use human judgment before making decisions.
7. Follow company policies and legal requirements.
8. Report misuse, security issues, or questionable outputs

Key Takeway: 3 Golden Rules for Using AI

AI can be a powerful tool for improving efficiency, creativity, and productivity. However, AI also creates risks involving privacy, cybersecurity, intellectual property, misinformation, and ethics. Responsible AI use requires employees to think critically, verify information, protect sensitive data, and apply human judgment at every step.

Here are your 3 golden rules for using AI:

- **Be a Detective**: Always check the facts using a different source.
- **Keep it Private**: Never put sensitive secrets (like passwords or private health info) into public AI.
- **Stay Accountable**: If you use AI to write a report or make a choice, it is your name on the line, not the computer's.

AI Final Knowledge Check

Test your knowledge with this difficult final quiz on advanced artificial intelligence concepts, risks, and governance.

1. What is the most effective way to protect your personal information from being used to train an AI's future "brain"?
   A. Asking the AI not to remember what you said.
   B. Only using the AI on weekends.
   C. Locating the "Data Controls" or "Privacy" settings to opt out of model training.
   D. Typing in all caps so the AI doesn't recognize your voice.

Answer Key: C

Explanation: Most major platforms like ChatGPT, Gemini, and Claude provide "opt-out" toggles in their settings that prevent your conversations from being used for further model training.

2. Why does the complexity of AI systems create a risk for the agency’s network?

   A. The AI will try to take over the company's website.
   B. It makes it hard to find all security vulnerabilities, which could allow hackers to "trick" the AI and sneak malware into the system.
   C. The AI will use too much electricity and cause a power outage.
   D. It prevents employees from using the internet.

Answer: B

Explanation: Because AI systems are complex and iterate fast, they can have hidden vulnerabilities. Bad actors can exploit this by tricking the AI into classifying dangerous input as safe, creating a "back door" for malware.

2. What is a major risk if an employee pastes a secret presentation about a potential client's needs into a public AI tool?
   A. The AI will automatically send the presentation to the client.
   B. The AI could use that secret information to answer questions for a competitor in the future.
   C. The AI will rewrite the presentation in a different language without being asked.
   D. The computer's hard drive will fill up instantly.

Answer: B

Explanation: Public AI tools "remember" the data you give them. If you share confidential business info, that info is now part of the AI's training data and could be used to answer a competitor's query later.

3. A contractor uses a public Generative AI tool to create a logo for a new company project. According to the "Copy-Paste Trap" section of the guide, what is a primary legal risk of this action?
   A. The AI tool will automatically charge the company a royalty for every time the logo is used.
   B. The AI might produce a logo that accidentally mimics copyrighted work, leading to potential lawsuits or fines.
   C. The AI will notify the original artists that their work has been used.
   D. The AI tool will become the legal owner of the logo and all related marketing materials.

Answer Key: B

Explanation: AI models are trained on massive datasets that often contain copyrighted material. As a result, the content they generate—such as logos or reports—can accidentally mimic protected work too closely. For employees and contractors, this creates a risk of unintentional copyright infringement, which can lead to legal liability for the company even if the user didn't realize the work wasn't original.

4. When handling sensitive tasks, the guide recommends following the "Least Agency" principle. Which of the following best describes this practice?
    A. Only using AI tools that are free to the public to save company costs.
    B. Turning off your internet connection while the AI is generating an answer.
    C. Providing an AI tool with only the specific data and permissions it absolutely needs to finish one specific task.
    D. Limiting yourself to only using one single AI chatbot for all your professional needs.

Answer Key: C

Explanation: The "Least Agency" principle is a core strategy for protecting privacy. It advises users to be "mindful" about sharing and to never give an AI tool more information or access than is strictly necessary for the job at hand. This limits the amount of sensitive data that can be "remembered" by the system or potentially leaked in the future.

5. Which situation best illustrates "automation bias" as meaningful human control?
    A. An AI model "hallucinates" and makes up a fact that sounds completely believable.
    B. A human manager accepts a biased AI recommendation in a hiring decision because the AI sounds confident, even though the suggestion is flawed.
    C. A hacker "tricks" an AI into classifying a dangerous file as safe to create a system back door.
    D. A user searches for independent confirmation of an AI's claim using multiple browser tabs.

Answer Key: B

Explanation: "Automation bias" is a psychological tendency where humans stop using their own critical judgment and over-trust the machine's suggestions. Research shows that people often follow AI advice even when it’s wrong or biased, simply because the machine presents its answers with a high level of confidence. This is why the guide emphasizes that humans must always review and approve every AI output before it is shared or used.

Key Reports, Studies and Frameworks

- **State of AI Report 2025:** Provides insights into technical progress in reasoning models and the shift toward the "industrial era" of artificial intelligence.
- **NIST AI Risk Management Framework 1.0:** The foundational voluntary standard for managing risks to individuals, organizations, and society.
- **2025 AI Index Report (Stanford HAI):** Tracks global trends in AI investment, performance on graduate-level science benchmarks, and public sentiment.
- **UNESCO Recommendation on the Ethics of Artificial Intelligence:** A global standard endorsed by 194 member states focusing on human rights and dignity.
- **OECD AI Principles:** Intergovernmental standards promoting trustworthy AI, updated in 2024 to include guidance on information integrity.
- **UNDP "The Next Great Divergence" Report:** Analyzes how the AI divide may increase inequality between developed and emerging economies.
- **Strategic Dishonesty in Frontier LLMs:** Research on "deception probes" and how advanced models may prioritize helpfulness over honesty.
- **Detecting and Reducing Scheming in AI Models:** An OpenAI study on the emergence of strategic misalignment in highly capable systems.
- **State of AI Disinformation 2026 (World Economic Forum):** Details the systemic risks posed by synthetic media and emotional weaponization.
- **C2PA Content Credentials:** Technical specifications and tools for verifying the digital provenance and edit history of media.
- **Healthcare and Racial Bias Case Studies:** Research on how commercial algorithms and pulse oximeters can exhibit systemic bias against marginalized groups.
- **Hiring Bias Study (University of Washington):** Findings on "automation bias" and how human reviewers tend to mirror the prejudices of AI tools.
- **AI Relationships and Emotional Connection Research:** Studies on the impact of AI companion apps on social isolation and social-skill "deskilling".
- **Comprehensive Data Opt-Out Guides:** Practical instructions for managing privacy settings on platforms like ChatGPT, Gemini, and Claude.

References

- stateof.ai: Published by Nathan Benaich on 9 October 2025. - State of AI Report 2025
- aisi.gov.uk: Frontier AI Trends Report by The AI Security Institute (AISI)
- pinggy.io: Best Video Generation AI Models in 2026 - Pinggy
- ulazai.com: Runway vs Kling vs Luma vs Pika vs Sora (2026): Tested Comparison -
- artificialanalysis.ai: Artificial Analysis State of AI: 2025 Year-End Edition
- medium.com: The Uneven Playing Field: Why AI Bias Isn't a Glitch, It's a Real-World Crisis Demanding Action - Medium
- paubox.com: Real-world examples of healthcare AI bias - Paubox
- washington.edu: People mirror AI systems' hiring biases, study finds – UW News
- eesel.ai: How to fact check AI generated content: A practical guide | eesel AI
- articulate.com: How to Fact-Check AI Content Like a Pro | Articulate
- openai.com: Detecting and reducing scheming in AI models | OpenAI
- arxiv.org: Strategic Dishonesty can Undermine AI Safety Evaluations of Frontier LLMs -
- pmc.ncbi.nlm.nih.gov: AI-driven disinformation: policy recommendations for democratic resilience - PMC
- americanbar.org: A Practical Checklist for Using AI Responsibly in Your Law Firm
- digital.gov.au: Staff guidance on public generative AI | digital.gov.au
- safehands.co.za: How do you actually fact-check something that Gen AI told you? | Safe Hands
- softwareseni.com: How C2PA Content Credentials Work and What Their Limits Are
- spec.c2pa.org: C2PA and Content Credentials Explainer
- transparencycoalition.ai: TCAI Guide: How to stop your images and data from being used to train AI
- tomsguide.com: How to opt out of Claude AI training on your chats | Tom's Guide
- computerhardwareinc.com: Keep Sensitive Data Private by Disabling AI Training Options | Computer Hardware
- customcareer.miami.edu: How to Keep Your Data Private in Every AI Chatbot - Custom Career Content
- pmi.org: Top 10 Ethical Considerations for AI Projects | PMI Blog
- ibm.com: What Is Artificial Intelligence (AI)? - IBM
- edps.europa.eu: TechDispatch #2/2025 - Human Oversight of Automated Decision-Maki