Session 2 Handout: Checking the Work and Protecting the Data

FAST-01 Applied AI Literacy at Work · 7formations · Instructor: Wadood Chaudhary · All examples fictional

Why this session matters

Two habits separate safe AI users from risky ones: they check output before it leaves their desk, and they never put restricted data into a prompt. This session turns both habits into routines you can run in under two minutes: the four-step check and the three-pile data sort.

Key concepts

1. Fluent errors hide in plain sight

In the spot-the-error drill you saw AI answers about Pioneer Valley Mutual Insurance Agency (fictional, Springfield) that read perfectly well and were wrong. The most dangerous errors are not typos. They are a wrong number inside a correct sentence, a rule turned upside down ("covered" instead of "excluded"), or a promise nobody approved ("we'll cover your hotel right away"). Nothing about the writing style signals these errors, so the only reliable defense is comparing the output with the source.

2. The four-step check: source, facts, numbers, tone

Run these four steps in order on anything AI drafts that someone else will read or rely on.

Step	Question	How to do it
Source	Where did this come from, and can I see it?	If you didn't supply a source and the answer contains facts, find the source now or treat the facts as unverified
Facts	Do names, dates, rules, and claims match the source?	Put the source next to the draft; check each claim
Numbers	Is every figure right?	Recalculate totals and percentages; trace each number to its origin
Tone	Is it right for the reader and the company?	Read it as the customer will; remove promises you can't make

Then record a decision:

- **Keep** — facts check out; light edits only.
- **Fix** — specific, fixable errors; correct them and recheck.
- **Discard** — wrong job, errors throughout, or produced from restricted data. Start again with a better prompt.

Asking the AI "Are you sure?" is not a check. Its answer is more generated text.

3. Data that never goes in

Massachusetts has one of the oldest data-security laws in the country. Under **G.L. c. 93H** and **201 CMR 17.00**, *personal information* means a Massachusetts resident's first name (or first initial) and last name combined with any of:

- a Social Security number;
- a driver's license or state-issued ID number;
- a financial account number, or a credit or debit card number.

Businesses that hold this data must protect it under a written information security program, train employees, and notify the Attorney General and affected residents after certain breaches. Pasting it into an unapproved tool can be exactly the kind of unauthorized disclosure the law is about.

Two more categories stay out of prompts:

- **Protected health information (PHI):** health information linked to a person (a patient's name with a visit reason, for example). Only allowed if the exact tool is covered by your employer's HIPAA business associate agreement *and* policy allows it. Class labs never use PHI.
- **Confidential business data:** pricing, bids, unreleased financials, trade secrets, source code, client files, HR or disciplinary matters, passwords, and API keys.

In April 2024 the Massachusetts Attorney General also issued an advisory confirming that existing consumer-protection, anti-discrimination, and data-security laws apply to AI. A wrong coverage statement or an unfair automated decision is still the business's responsibility.

4. Three data piles: OK, remove first, never

Pile	Examples	Action
OK	Published hours, approved FAQ, public fact sheets, blank templates, your own non-sensitive notes	Use in the approved tool
Remove first	Notes with a contact's phone number; a complaint email with a customer's name and order number; a whiteboard photo with a phone number	Replace with placeholders ([CUSTOMER], [ORDER#]) or crop; then use
Never	Name + SSN, license, or account/card number; PHI; passwords; bids; HR files; source code	Do not enter. Use a different method

5. Your employer's policy

Every employer's AI policy should answer four questions. If yours doesn't, ask your manager.

1. Which tools are approved, and with which accounts?
2. What data may each tool see?
3. Who reviews AI output before it goes out?
4. How do I report a mistake or a data slip?

If your employer has no policy yet, the course uses a model policy based on these rules. Approval matters because enterprise versions of tools (Copilot with a work account, ChatGPT Business or Enterprise, Gemini in Workspace, Claude for Work) come with contractual data terms, such as not using your prompts to train public models, that personal accounts may not have. Even so, assume prompts and outputs can be logged and reviewed.

Worked example: checking a coverage reply

Source (fictional PVM Handbook 4.2): water backup is excluded unless endorsement WB-1 is purchased; WB-1 pays $5,000 per occurrence after a $500 deductible; report claims within 30 days.

AI draft: "Your policy covers sewer backup up to $10,000. Report within 60 days."

Step	Finding
Source	Handbook 4.2 supplied; can compare
Facts	"Covers" contradicts "excluded unless WB-1"
Numbers	$10,000 vs $5,000; 60 vs 30 days; deductible missing
Tone	Neutral, but implies a promise of coverage
Decision	**Discard.** Rerun with the excerpt pasted and "use only this text; say 'not in source' if missing"

Worked example: removing before using

You want AI to summarize your notes from a vendor call. The notes include "Called Dana R. at 508-555-0199 about the late resin delivery."

Sanitized: "Called [VENDOR CONTACT] about the late resin delivery." The summary still works, and nothing personal goes in.

Checklist: the two-minute pre-send check

- ☐ I can see the source the facts came from.
- ☐ Every name, date, rule, and claim matches the source.
- ☐ Every number is recalculated or traced.
- ☐ The tone fits the reader, and there are no promises I can't keep.
- ☐ I recorded keep, fix, or discard.
- ☐ Nothing in my prompt was in the "never" pile.

Do and don't

Do	Don't
Put the source next to the draft and compare claim by claim	Judge accuracy by how well it reads
Recalculate numbers yourself	Ask the AI to check its own numbers
Replace names and numbers with placeholders	Paste raw emails, spreadsheets, or screenshots
Report a data slip the same day	Ask the tool to "forget" it, or delete the chat and say nothing
Use the approved tool with a work account	Use a personal account because it's quicker

Try this at work

1. Take one AI draft you (or a coworker) produced recently on non-sensitive content. Run the four steps and record a decision.
2. Make a list of five data items in your role: two "never," two "remove first," one "OK." Check them with your manager.
3. Find your employer's AI policy and answer the four policy questions on one card.

Glossary

Term	Plain-language meaning
Four-step check	Source, facts, numbers, tone, then keep / fix / discard
Personal information (Mass.)	Name plus SSN, license/state ID number, or financial account/card number (G.L. c. 93H; 201 CMR 17.00)
PHI	Health information that identifies a person (HIPAA)
WISP	Written information security program required by 201 CMR 17.00
Placeholder	A label like [CUSTOMER] that stands in for removed details
Data slip	Restricted data entered where it shouldn't be; report it
Enterprise data protection	Contract and technical terms in work versions of AI tools governing how prompts are stored and used

Key takeaways

- Check against the source; style is not evidence.
- Numbers get recalculated, not trusted.
- Never: name + SSN/license/account number, PHI, confidential business data, passwords.

This handout is training material, not legal advice. Follow your employer's policy and counsel. Facts as of October 2026.