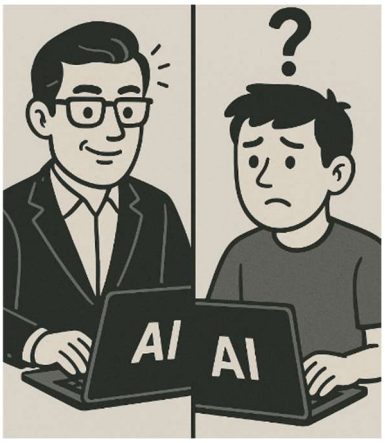
EVALUATING GENERAL AI USE

Using AI "Successfully"



In general, your **"success" when *attempting* to use AI** to complete a task can vary greatly depending on a great many things including the following.

- the **nature of the task** (how good is AI at performing it)
- your ***ability* to know** if the task has been completed in a "**successful"** way
- ***whether*** you are actually **checking** if the task has been completed successfully
- how ***thoroughly*** and the ***process* you're using to check** if the task has been completed successfully.

Basic Metrics of AI "Success" (= General ML Analysis Success)

Just like in any machine learning analysis, the "success" of using AI in a task should obviously be evaluated on things like:

- **accuracy** and
- **efficiency** (ie. the main reason people use AI)

Other Important Metrics of AI "Success" (= General Data Science Analysis Success)

However, in general, just like when you complete any type of data science analysis, there are **other metrics of success** that you want to be thinking about as well including the following.

- **Goal Alignment:** Is your work in **alignment with broader research/organizational goals**?
- **Trust/Credibility:** Your report's ability to garner **trust/credibility** in:
  - *your* work and
  - the *data science community* in general
- **Transparency:** How much do you (and your reader) understand all the steps and decisions taken in the pursuit of your research goal, and the potential pros/cons of each of the decisions that were made?

- **Reproducibility:** How easy would it be for you (and your reader) to reproduce your findings?
- **Robustness**: Are you seeking a thorough, expansive understanding of how to go about answering your question or pursuing your goal? How likely is it that other data scientists pursuing this same goal/question (also in a robust way) will arrive at your same conclusion?
- **Responsibility**: Do you understand the models/methods that you are using and when they should and should not be used? Have you checked the appropriate conditions?
- **Well-Motivated:** Is your work/decisions **well-motivated** such that interested parties would know **why** your work is useful to them and what their **next steps** might be based on the result of your work?
- **Communication:** Are your insights **communicated well?**
    - *Important Note: Just because your work is written with good grammar in a snappy, persuasive, emotionally evocative tone does not mean that you have communicated your research findings well. See all of the items above for how to improve the communication of your findings in a meaningful way.*

DATA SCIENCE: EXAMPLE OF A TASK YOU SHOULD NOT USE AI FOR… YET.

AI Tool *Correctness* Warning

GENERATING data science **code** and **conceptual/interpretation content** with AI tools like ChatGPT still gets many things wrong.

Professional Penalties for Irresponsible AI use in Machine Learning

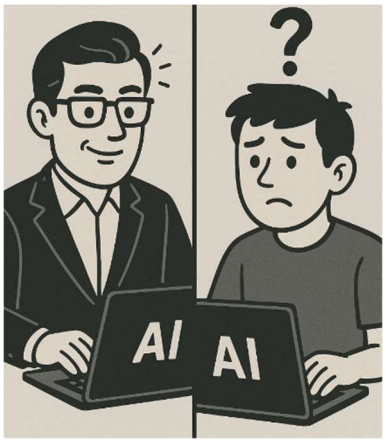
Using AI for data science, without a **deep understanding of the topic** makes it *very likely* that you will:

- **Miss valuable insights** via **flawed/incorrect analyses**
- Produce **misleading/incorrect interpretations**
- **Erode trust** in:
  - *Your* work
  - The data science research community in general

Developing the Expertise to CATCH Data Science AI Errors

By engaging with ***RELIABLE* data science resources** in this class (ie. NOT AI) and developing more of a fluid, internalized intuition for these topics you will develop more of an expertise to **catch data science** related **AI errors.**

Thus, if you choose to use AI for data science in the future, you will be better able to catch these errors.



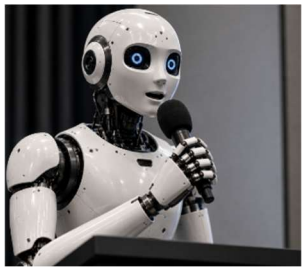
Short-Term AI Efficiency ≠ *Long-Term* Efficiency

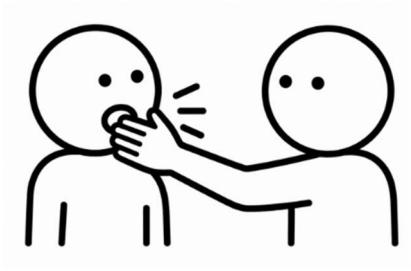
If your code/work is full of bugs and inaccuracies (especially due to AI errors) in your job, the amount of time it takes to go back and fix these errors usually does not make unchecked/irresponsible AI use worth it in the end.

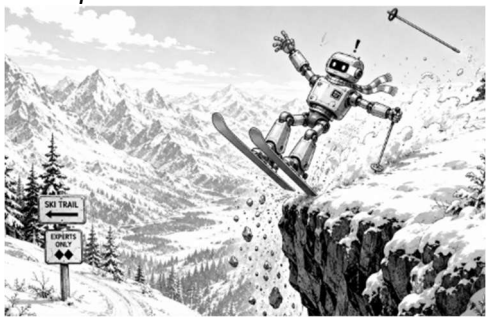
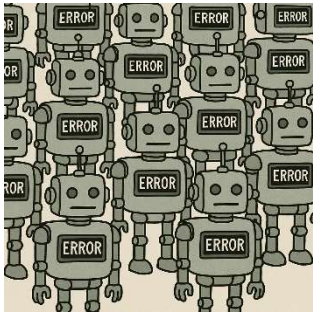
AI + Data Science *Concepts/Interpretations*

AI tools like ChatGPT make ***A LOT of incorrect claims/interpretations about data science concepts.***

Common *Conceptual* Data Science AI Slop

Type	Professional Penalty	Course Penalty
"ChatGPT linguistic style" (ANY content written in this style) *Ex: Too many emdashes.* **What counts as "ChatGPT linguistic style"?** (Just like in your future career, it is incumbent upon *you* to know and keep up to date with what people think "looks" like "ChatGPT linguistic style" if you plan to use it.) 	• Even if characteristic "ChatGPT linguistic style" is used for non-technical content, reader **loses *some* trust** that you actually understood and/or looked over the technical work. • Potentially not the appropriate **tone/audience** for the analysis. • Overhyped tone can be **misleading.**	• **-1.5 points** additionally deducted from assignment.
"Provable" Unchecked AI Content Generation • *Ex: Leaving the AI prompt/response/instructions in a copy-paste "Sure, here is a k-means elbow plot below".* • *Ex: A hallucinated written response that is completely unrelated/misaligned with what other code/elements of your assignment says.* 	• Reader **loses *pretty much all* trust** that you actually understood and/or looked over the technical work. The work could be full of errors, for all they know. Reading and **using the insights** from this work becomes very **risky for the reader**, if they do not wish to have a flawed understanding of the topic.	• **0 on the assignment** ("drop the 2 lowest" doesn't count). • The assignment counts as a **"AI Slop Warning"** (see tier system) • ***Repeated*** "Provable" Unchecked AI Content Generation of this nature is subject to an **academic integrity violation.**

<u>"Blindfolded Tour Guide"</u> Speaking in vague, speculative generalities about the POTENTIAL nature of your dataset (or some plot) that you have ACTUALLY learned (or created) somewhere else in your analysis.  *<u>Ex</u>: Creating a plot, but when "interpreting" the plot saying something like…* ○ *"Plots like this MAY say something like…", or* ○ *"Plots like this are interpreted like….".* ○ *Ie. Not interpreting YOUR actual plot, but talking in vague generalities.*	• You didn't actually interpret YOUR plot. Without an interpretation of YOUR results, the **analysis is incomplete** and **uninformative** in this place.	• **-1.5 points** additionally deducted from assignment. • **0 points** on all **questions** related to the **"blindfolded" interpretation.** • The assignment counts as a **"AI Slop Warning"** (see tier system) • Your response may be used as an **AI slop example** in the next assignment.
<u>"TMI (Too Much Information) of AI Slop"</u> *Answering a question/prompt correctly. But then giving **additional information** that was not asked for, that is **incorrect, overly speculative, illogical,** or **superfluous**.* *<u>Ex</u>:* *<u>Question</u>: How many clusters does the elbow plot suggest this dataset has?* *<u>Answer</u>: "The elbow plot suggests there are 3 clusters, because this is where the elbow is. ~~This is probably because k-Means tends to work great for sports datasets".~~* 	• While your initial interpretation might be correct, AI likes to inject **speculative, incorrect, illogical hallucinations to try to explain WHY** a particular result was found. • First, this potentially injects **unnecessary errors** into your analysis. • Second, your algorithm <u>interpretations *could* always be incorrect</u> (no technique makes perfect interpretations for ALL datasets). However, a fake explanation added on gives your potentially fallible interpretation a **false sense of confidence** which is **misleading.**	• **-1.5 points** additionally deducted from assignment. • **0 points** on the entire **question with the TMI.** • The assignment counts as a **"AI Slop Warning"** (see tier system) • Your response may be used as an **AI slop example** in the next assignment.

<u>Out of Scope Errors</u> *An **out-of-scope error** in this class counts as an incorrect or illogical claim about some data science-related concept that was not learned in this class.*  	• ChatGPT can produce **A LOT of errors** into even a short sentence about data science conceptual topics. **You should be skeptical of AI generated claims.**	• **-1.5 points** additionally deducted from assignment. • **0 points** on the entire **question with the out-of-scope error.** • The assignment counts as a **"AI Slop Warning".** • Your response may be used as an **AI slop example** in the next assignment.
<u>AI Conformity Errors</u> *Some AI content/code generators will display the **same general idiosyncratic structure/logic** for quite a few of the assignment questions and prompts that I might ask you in this class. I usually run assignment/project questions through AI tools to see what kind of idiosyncratic incorrect/illogical/misaligned answers it produces ahead of time.* *If your answer displays one of these general idiosyncratic structures and it is **wrong**, **illogical**, or **not in alignment with the broader research goals** of the analysis, then this will also be penalized for AI slop.*  	• Promoting and disseminating AI errors ***individually*** in an analysis is unideal on its own. • However, when ***multiple people*** end up promoting and disseminating the ***same AI error***, more people are likely to believe this error, and it may become **harder for *truthful, accurate analyses* to be believed.**	• **-1.5 points** additionally deducted from assignment. • **0 points** on the entire **AI conformity error.** • The assignment counts as a **"AI Slop Warning"** (see tier system) • Your response may be used as an **AI slop example** in the next assignment.

AI + Data Science *Code*

When **writing code** for data science analyses, AI tools like ChatGPT can insert random **parameter changes** *(different from the code given in class)* or **other subtle changes** that are:

- insensible for the broader analysis or how the technique is supposed to be used,
- not what I'm asking for and/or
- unexplained.

that can subtly *derail the whole analysis* via:

- **missing useful insights** or
- creating **misleading interpretations.**

Common *Code* Data Science AI Slop

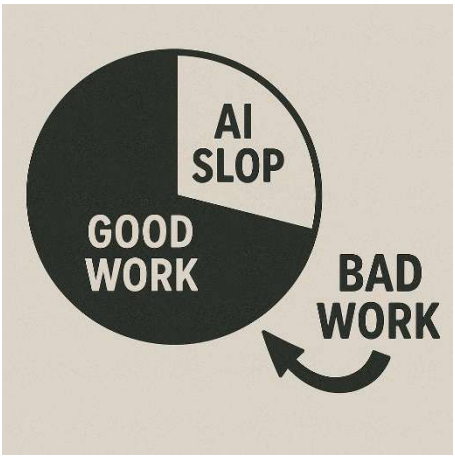
Type	Professional Penalty	Course Penalty
Unexplained, "Off-Menu" Code Errors Using code that creates **different results** than the code discussed in class, and the change either: • **is incorrect/not sensible** for the broader analysis; • is **not explained;** or • is explained, but the **reason is not sensible.** *Ex: The particular code in class uses **Euclidean distance**, but your code randomly decides to use the **cosine similarity** for a given technique. No explanation as to why is given.*	• Subtle code/parameter changes to analyses without thinking about the broader effects of your choice can **derail a whole analysis.** • We'll examine examples of this in class.	• **-1.5 points** additionally deducted from assignment. • **0 points** on the entire **question.** • The assignment counts as a **"AI Slop Warning"** (see tier system) • Your response may be used as an **AI slop example** in the next assignment.

Note: You will only lose -1.5 points for each distinct *type* of AI slop once per assignment. So for instance if you have the following in an assignment, then you will lose just (*-3 points)* on this assignment, *in addition to any points lost for incorrect answers.*

- 4 distinct questions with "TMIs of AI slop" (-1.5 points)
- "ChatGPT linguistic style" (-1.5 points)

AI SLOP IN THE GROUP PROJECT

Professional Penalties for *Consistent* AI Slop



In your future career, if your work reflects ***consistent* AI slop** that you should have reasonably anticipated, your colleagues may be **less likely to want to collaborate with you** on shared projects that they are hoping to use to advance their career aspirations.

Project Group Tiers

At the end of the semester, you will work in groups of 3-4 to complete the final project. To mirror these professional penalties, at the end of the semester, students will end up in one of two tiers.

- **Tier 1 Students:** *At most 2 assignments* in which there was an "AI slop warning".
- **Tier 0 Students:** *More than 2 assignments* in which there was an "AI slop warning".

Forming Project Groups Based on Tiers

- A group of ALL **tier 1** students can work together.
- A group of ALL **tier 0** students can work together.
- In order for a mixed group with **tier 1** AND **tier 0** students to work together, **students in the proposed group must email me that they approve of this.**

AI Slop Penalties in the Group Project

In order to ensure fairness in group contributions, if a group member significantly lowers the project grade due to **excessive AI slop** (or not completing the prompts in their assigned section):

- this **teammate will be graded exclusively on their contributions** and receive that as their final project grade
- the remaining teammates will be graded on the collective grade (minus this teammate's contributions).