**1) Common AI & Human Mistakes**

These errors occur **frequently** in both **AI-generated** and **human-written** code, often due to a lack of contextual understanding or fundamental programming habits.

**Common Mistakes:**

✅ **Incorrect Handling of Missing Values (AI & Humans)**

* **Example:** Dropping all NA/NaN values without checking their impact.
* **Why?** Both AI and humans often **default to dropping missing data** instead of **imputing** or **analyzing** why data is missing.

✅ **Inefficient Use of Loops Instead of Vectorized Operations (AI & Humans)**

* **Example:** Iterating through a Pandas DataFrame or R DataFrame row-by-row instead of using built-in vectorized functions like .apply() or dplyr::mutate().
* **Why?** Many beginners and AI-generated scripts **default to loops** instead of **optimized operations**.

✅ **Forgetting to Set index=False or row.names = FALSE When Saving Data**

* **Example:** Writing CSV files with an extra unwanted index column.
* **Why?** AI often forgets this small detail, and humans unfamiliar with Pandas or R’s write.csv() function do too.

✅ **Incorrect Type Conversions Without Error Handling**

* **Example:** Converting a column to numeric without handling errors, leading to NA or NaN values.
* **Why?** Both AI and humans forget that not all values may be convertible (e.g., "Tree\_Height\_m" might have unexpected non-numeric values).

✅ **Forgetting Edge Cases in Conditional Statements**

* **Example:** Creating if-else conditions for classifying data but missing an edge case.
* **Why?** AI often **mimics surface-level logic** without testing edge cases, and humans tend to **overlook rare cases**.

**2) Common AI Mistakes (Less Common for Humans)**

These are **errors AI frequently makes** but **humans tend to avoid** due to experience, intuition, and logical reasoning.

🔴 **Overcomplicating Simple Calculations**

* **Example:** Applying min-max normalization to pH values (df['pH'] = df['pH'] / df['pH'].max()) when pH is already on a fixed scale (0-14).
* **Why?** AI tries to **apply standard transformations to everything**, even when unnecessary.

🔴 **Forgetting errors='coerce' in Pandas Date Conversions**

* **Example:** AI-generated code often assumes all dates are formatted correctly and does not account for parsing failures.
* **Why?** Humans usually debug parsing errors immediately, while AI lacks context awareness.

🔴 **Over-reliance on Default Parameters Without Checking Results**

* **Example:** Using default scaling methods or classification thresholds without verifying if they make sense.
* **Why?** AI assumes default values are always correct, while humans tend to **double-check logic**.

**3) Rare AI & Human Mistakes**

These are **unusual errors** that happen only in **specific cases**, either due to misinterpretation of the problem or a highly specific edge case.

🟡 **Incorrect Order of Operations in Data Cleaning Pipelines**

* **Example:** Dropping missing values **before** attempting to convert them instead of **after** handling them properly.
* **Why?** Rare because most humans naturally check data before removing parts of it, and AI usually follows standard cleaning steps.

🟡 **Incorrect Statistical Transformations Applied to the Wrong Features**

* **Example:** Applying log() transformations to categorical variables.
* **Why?** AI occasionally misapplies statistical concepts, and humans rarely make this mistake unless unfamiliar with the dataset.

🟡 **Misusing Boolean Conditions Leading to Logical Errors**

* **Example:** Using if x > 5 and x < 3: in Python (which is always False).
* **Why?** AI-generated code can misinterpret numeric conditions, but experienced programmers **rarely** write logically impossible conditions.

**How Do I Know What Is Common or Rare?**

✅ **Trained AI Patterns:**  
I’ve seen **many examples of AI-generated code**, and I recognize **patterns of AI mistakes**, especially in data science workflows.

✅ **Human Programming Experience:**  
I base these classifications on **real-world coding challenges** that students, early-career data scientists, and professional programmers frequently encounter.

✅ **Research on AI Code Generation Errors:**  
I reference **published studies** on LLM-generated code issues, where AI is tested against real-world datasets.

**TL;DR: Key Differences Between AI & Human Errors**

| **Type of Error** | **Common for AI?** | **Common for Humans?** | **Why?** |
| --- | --- | --- | --- |
| Forgetting index=False in Pandas | ✅ Yes | ✅ Yes | Both AI & humans overlook it. |
| Using inefficient loops instead of vectorized ops | ✅ Yes | ✅ Yes | Common beginner mistake. |
| Dropping all missing values without checking impact | ✅ Yes | ✅ Yes | Both AI & humans default to this. |
| Over-normalizing data unnecessarily | ✅ Yes | ❌ No | AI applies transformations blindly. |
| Forgetting to handle parsing errors (errors='coerce') | ✅ Yes | ❌ No | AI assumes data is always clean. |
| Applying statistical methods to categorical data | ✅ Yes | ❌ No | AI misapplies transformations. |
| Using impossible logical conditions | ✅ Yes | ❌ No | AI sometimes misinterprets logic. |

**Final Thoughts**

* **AI makes systematic, repeatable mistakes** due to **lack of contextual understanding**.
* **Humans tend to make mistakes due to oversight** rather than fundamental misunderstanding.
* **Combining AI-generated code with structured debugging** (like what we’ve built in this course) is the best way to **help students recognize and correct mistakes efficiently**.