# Checkpoint 5 Guide: Presentation and Communication

## Principles

* **Tell a story, not a workflow dump**
  + The audience doesn’t need every preprocessing detail. Focus on what the problem is, why it matters, and what you learned.
* **Transparency**
  + Be upfront about limitations, assumptions, and failure cases.
  + Highlight what the model *can* and *cannot* do.
* **Reproducibility**
  + Slides and reports should match what’s in the repo (no cherry-picked screenshots or manual edits that can’t be reproduced).
* **Audience awareness**
  + Tailor language and depth depending on whether you’re presenting to technical peers, domain experts, or decision-makers.

## Presentation Content

* **Problem Recap**
  + Stakeholders, decision context, unit of analysis, KPIs.
* **Data Story**
  + Where the data came from, main limitations, key EDA insights.
* **Feature Engineering Story**
  + What features were kept, dropped, and created — and why.
* **Modeling Journey**
  + Baselines → more complex models → final choice.
  + What didn’t work and why it was discarded.
* **Final Results**
  + KPI performance (primary and secondary) with context.
  + Visualizations: ROC/PR curves, calibration plots, error distributions.
* **Limitations & Next Steps**
  + Where the model might fail.
  + Additional data or features that would improve it.
* **Business/Decision Relevance**
  + How the outputs connect to actual decisions and trade-offs.

## Tools and Formats

* **Slides**
  + Tools: PowerPoint, Google Slides, Keynote, or Jupyter-based slide decks (e.g. [RISE](https://rise.readthedocs.io/)).
* **Notebooks for reproducibility**
  + Should be runnable end-to-end to reproduce reported metrics and plots.
* **Interactive options (optional)**
  + Dashboards: [plotly-dash](https://dash.plotly.com/), [streamlit](https://streamlit.io/).
  + Reports: [jupyter-book](https://jupyterbook.org/), [quarto](https://quarto.org/).

## Deliverables

* **Written / Conceptual**
  + **Slide deck** (presentation.pdf or .pptx/Google Slides link) covering the elements above.
  + **Executive summary document** (summary.md or .pdf) with:
    - Problem, data sources, KPIs, results.
    - Final model choice with justification.
    - Limitations and next steps.
* **Code / Repo Artifacts**
  + **notebooks/final\_results.ipynb**:
    - Clean, readable notebook showing final model training, evaluation, and plots.
    - Must produce all metrics/visuals shown in the presentation.
  + **results/final/ folder**:
    - Saved KPI tables, plots (ROC/PR, calibration, error analysis).
  + **artifacts/final\_model.pkl** (or .joblib):
    - Serialized final pipeline/model.
  + **presentation/ folder**:
    - Slide deck source file (if not Google Slides) + exported PDF.
    - Executive summary (summary.md or .pdf).
  + **Reproducibility check** (Makefile or run.sh):
    - One command/script to regenerate final results from raw data → features → models → figures.

## Iteration and Communication Loops

* **Feedback cycles**
  + Practice presenting to peers before final delivery. Note where people get confused or skeptical.
* **Iterate on clarity**
  + If you can’t explain why you chose a model or dropped a feature in one slide, you probably don’t understand it well enough.
* **Traceability**
  + Every plot or number in the presentation must trace back to a notebook and dataset in the repo.