APPENDIX C | COMPANION TO CHAPTERS 1 AND 8

SOP: AI Bias Mitigation

A standard operating procedure for catching bias before it reaches a report.

Purpose. A repeatable procedure for identifying, documenting, and mitigating bias when AI tools support intelligence analysis.

Scope. Any AI-assisted task where model output informs a judgment, flags a person or entity, or ends up in a report.

Task / case	AI tool and version	Date

Analyst	Reviewer

PART 1 Pre-use screening

Complete before the tool touches a live task.

- ☐ Tool limitations and vendor disclosures reviewed
- ☐ Training data representativeness assessed (global names, languages, regions, platforms)
- ☐ Who labeled or interpreted the training data identified, if known
- ☐ Edge case and adversarial testing confirmed (slang, low-quality data, non-Western names), or flagged as unknown

Notes

PART 2 During use

Follow at the point of analysis.

- ☐ Prompts are neutral and well structured, nothing presumes the conclusion
- ☐ Input sources are diverse and balanced, not one language, region, or platform
- ☐ Output checked for stereotypes or familiar-threat narratives
- ☐ Clear trail for outputs: I can explain WHY an account was flagged or a person identified

Notes

PART 3 Post-output validation

Complete before any AI-assisted finding moves forward.

- ☐ Validated against an alternative source or tool, never one model as sole confirmation
- ☐ Human critical review completed: does the model consistently miss certain subsets (dialects, languages, faces, demographics)?
- ☐ I can override or challenge the output, it doesn't just pass through
- ☐ Tool use and approvals documented per my organization's requirements
- ☐ Classification and handling of the output confirmed

Notes

PART 4 Where did the bias get in?

Bias can enter at any stage of the AI lifecycle. When an output looks wrong, use this to trace it, then check the stage(s) you suspect.

Stage	Risk
Data collection	Data that doesn't represent the real world produces a model that inherits that skewed view.
Model design	How inputs and features are structured can hard-code assumptions about what is normal.
Optimization goals	A system built for speed or accuracy without fairness can learn shortcuts that harm certain groups.
Deployment	Real-world conditions differ from training conditions. Without monitoring, errors compound.

☐ Data collection ☐ Model design ☐ Optimization goals ☐ Deployment

PART 5 Escalation

If a bias concern could affect a live case, decision, or person:

1. Halt forward distribution of the affected output.
2. Document the concern and the lifecycle stage it likely came from (Part 4).
3. Escalate to a supervisor or designated review authority before the output is used in any operational decision.
4. Record the resolution and any corrective action.

Escalation log

Date	Concern	Stage	Escalated to	Resolution

Date	Concern	Stage	Escalated to	Resolution

PART 6 Quick bias mitigation checklist

The short version from Chapter 1, for when you're moving fast. Initial each step.

#	Task	Complete?	Initials
1	Review tool limitations and disclosures		
2	Use well-structured and neutral prompts		
3	Ensure input sources are diverse and balanced		
4	Validate AI outputs against alternative sources and tools		
5	Conduct a human review and critical analysis of outputs		
6	Document tool use and approvals		
7	Ensure proper classification and handling of outputs		

We can't eliminate bias altogether. We can shrink it a lot with design, oversight, and a documented procedure like this one. Human oversight is non-negotiable.