CCJ Council on Criminal Justice

Appendix C: System Complexity and Interpretability Assessment

AI systems vary in their complexity, transparency, and the difficulty of validating their behavior and impact. The characteristics highlighted here help determine how much scrutiny, testing, and monitoring are needed.

Dimensions to Assess

1. Decision Transparency

- **More Transparent:** The system follows clear, documented rules or formulas that can be traced step-by-step. Operators can explain how and why the system produced a specific output.
- **Less Transparent:** The system uses methods that are opaque or difficult to understand. The exact reasoning path may be unknown or not intuitive to human operators.

2. Predictability and Consistency

- **More Predictable:** Given the same inputs, the system will always produce the same output. Its behavior is stable and well-understood.
- **Less Predictable:** The system may produce novel or unexpected results. It might learn and adapt over time or exhibit behaviors that surprise operators.

3. Validation Difficulty

- **Easier to Validate:** The system's performance can be tested in a straightforward fashion. Errors are easy to identify, and the system's accuracy can be verified through standard quality assurance processes.
- **Harder to Validate:** The system's behavior is difficult to test comprehensively. It may process complex, ambiguous inputs (like images or natural language) where "correct" outputs are subjective or context-dependent. Unexpected behaviors may only emerge after deployment.

4. Adaptability

- **Static Systems:** The system's logic only changes when humans manually reprogram it. Its rules are fixed.
- **Adaptive Systems:** The system may learn from new data and change its behavior over time. This adaptability creates uncertainty about future performance.

Assessment Questions

Answer these questions to help understand your system's complexity profile:

1. Can you trace and explain each step the system takes to reach its outputs?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

2. Does the system always produce the same output for the same input?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

3. Does the system use only structured, well-defined data inputs (not images, free text, or audio)?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

4. Are you confident you can fully validate the system's performance before deployment?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

5. Is the system's logic fixed and static (it won't learn, adapt, or develop new behaviors over time)?

   ☐ Yes, completely → lower complexity ☐ Partially

   ☐ No or with great difficulty → higher complexity

How to Use This Assessment

Systems with **lower complexity** (transparent, predictable, easy to validate, static) should be subject to standard oversight approaches:

- Audit the rules and logic
- Verify calculations and outputs
- Use standard quality assurance processes

Systems with **higher complexity** (opaque, ambiguous, hard to validate, processing unstructured data) should be subject to enhanced oversight:

- Demand explainability capabilities and case-specific decision rationales
- Require independent technical validation
- Plan for ongoing performance validation post-deployment

Document your findings: Note in your **Classification Memo** the system's complexity profile and what additional scrutiny or safeguards this triggers.