

# Domando el Sycophancy en SLMs locales con Shift-Left Data Parsing

Cuadernillo | Vol. 1

mercedev.es — 2026-05-09 | Fase Épica 2 - Fase 3 (Orquestación  
de Contenidos)

## El Desafío (Síntoma)

---

Los SLMs locales como Qwen 2.5 Coder sufren de Sycophancy (sesgo de complacencia) y Checkbox Hallucination. Si les dices que no lean los pasos futuros, los leen y se los inventan para agradar.

## La Maniobra (Lógica)

---

Para domarlos, se implementó el Shift-Left Data Parsing. Se refactorizó el script `scripts/merci/merci-ssot.py` para aplicar filtrado Regex estricto antes de enviar el texto a la IA. Se usó la expresión regular `re.findall(r'\*\*Hecho:\*\*(.*?)\*(?=\*\*Detalle técnico:...)` para amputar matemáticamente los bloques de Contexto y Siguiendo Pasos, asegurando que solo se envíe el texto relevante.

## El Aprendizaje / Deuda Técnica

---

La solución consiste en purgar el contexto mediante código nativo (Python) antes de la inferencia. Esto garantiza un 0% de alucinaciones prospectivas y permite una mayor confianza en las instrucciones dadas a los SLMs. Se ha asumido una pequeña deuda técnica en términos de mantenimiento del script, pero el beneficio es significativo para la precisión y fiabilidad del sistema. ``