Honest Comparison: Java Migration Success Rates
(Different evaluation criteria - not directly comparable)

Ours (Sonnet 3.5)
Java 21
[Agent Self-Report]
⚠ Not Independently Verified — **67.5%**

Gemini 2.5
(FreshBrew)
Java 17
[Build+Test+Coverage] — **52.3%**

Gemini 2.5
(FreshBrew)
Java 21
[Build+Test+Coverage] — **49.8%**

SD-Feedback
(MigrationBench)
Java 17, Minimal
[Independent Eval] — **62.3%**

SD-Feedback
(MigrationBench)
Java 17, Maximal
[Independent Eval] — **27.3%**

Success Rate (%)

Legend:
- MigrationBench (Independent Eval)
- FreshBrew (Independent Eval)
- Ours (Agent Self-Report)

⚠ WARNING: These methods use different:
• Target Java versions (17 vs 21)
• Evaluation criteria
• Verification methods