Accuracy-Cost Trade-off (Gating Sweep)

F1 Score vs Verifier Invocation Rate (EscRate %)

Data points:
- [0.4, 0.6]
- [0.3, 0.7]
- [0.2, 0.8]
- [0.1, 0.9]
- [0.05, 0.95]
- [0.01, 0.99]