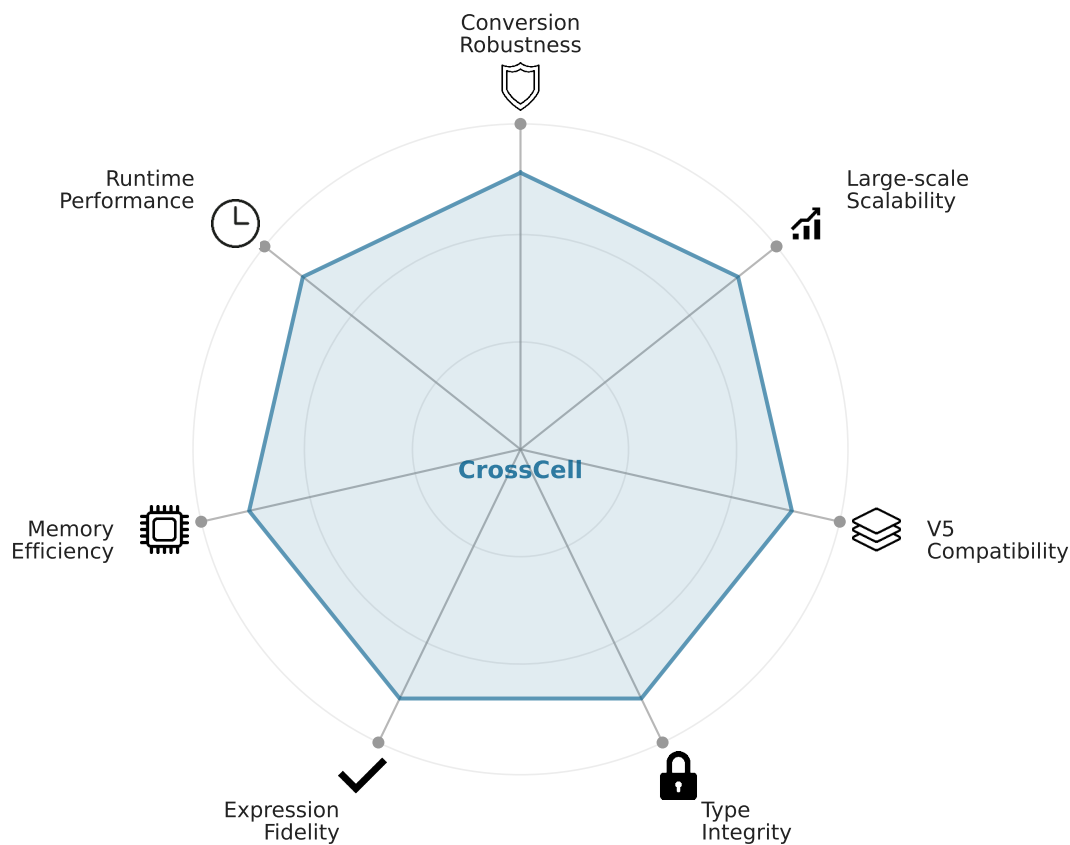


Seven-Dimension Evaluation Framework



Reproducibility

-  Version-locked environment
-  3× replicate runs per test
-  All code, data, notebooks public
-  65 datasets × 5 tools × 7 dim = 2,275 test points