

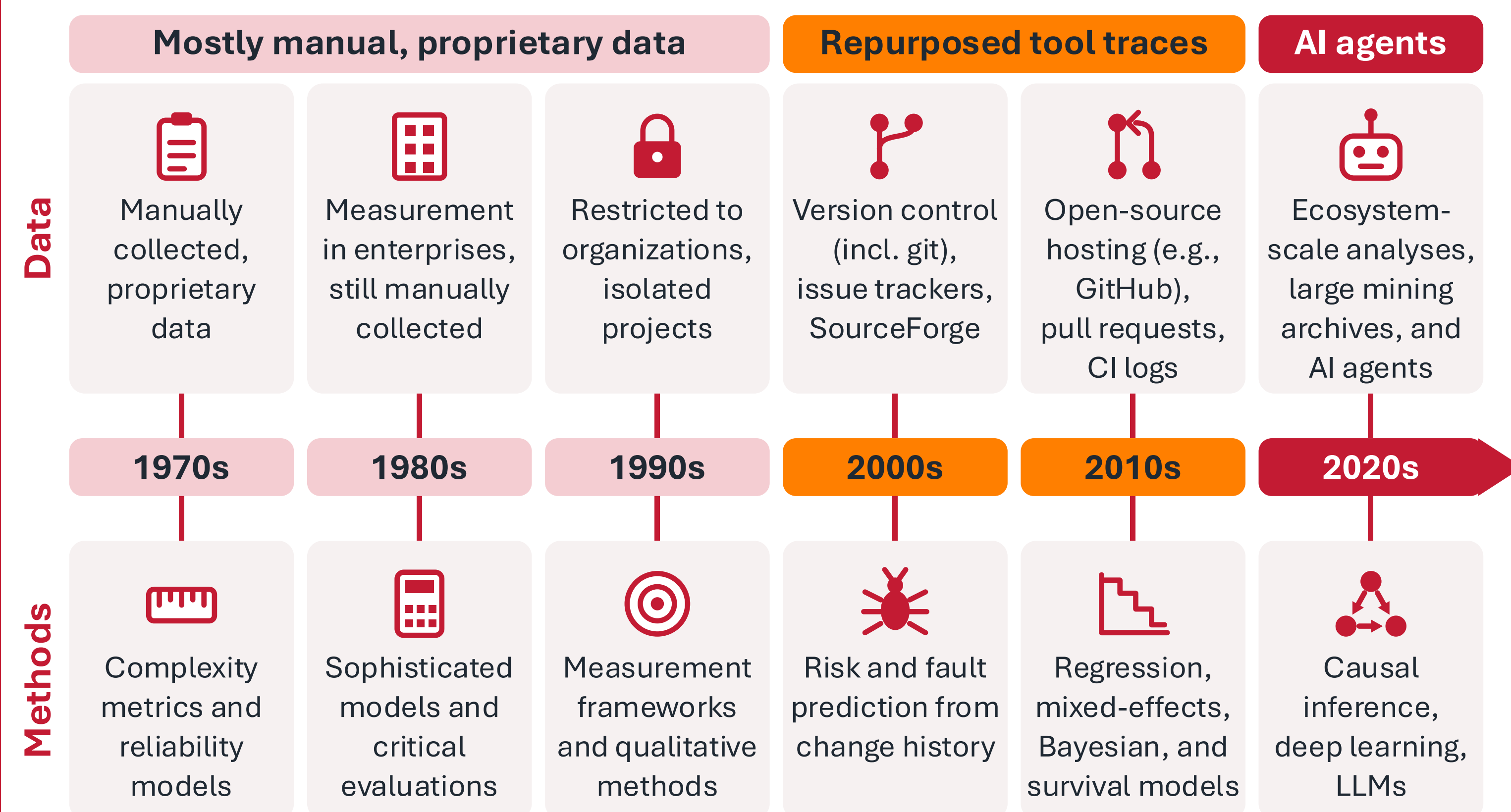
The Ground Is Shifting: A Reflection on the Foundations of Software Measurement

Thomas Bock, Audris Mockus, and Bogdan Vasilescu

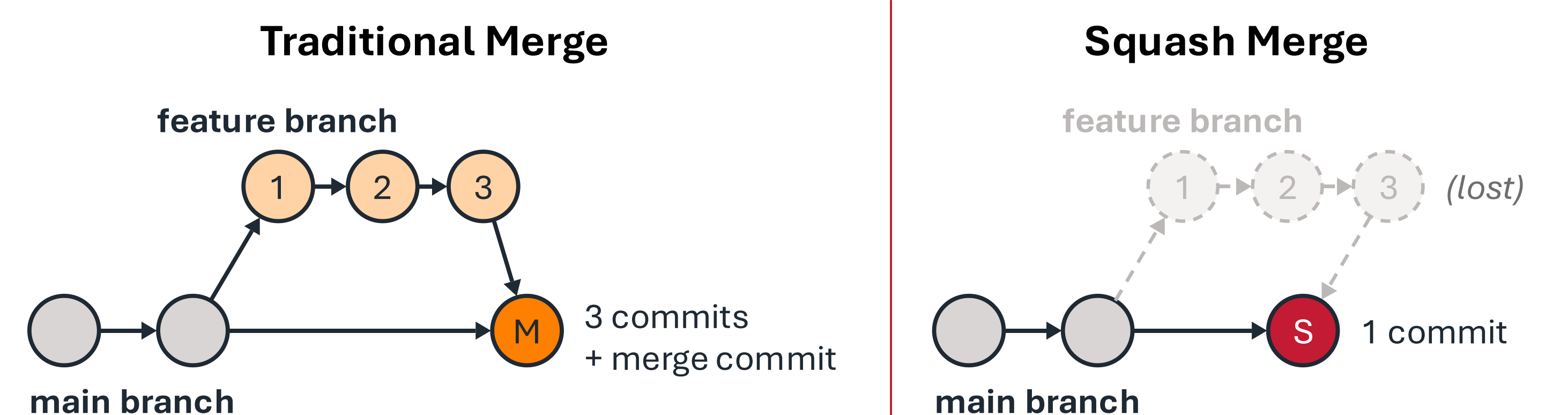
The History of Software Measurement

Data sources and analytical methods have evolved over six decades of empirical software engineering research.

Would a historic analysis repeated today without adjusting for changes in how the data are generated lead to the same conclusions?



Squash-Merging Hides Developer Activity



The magnitude of developer activity hidden by squash-merging:

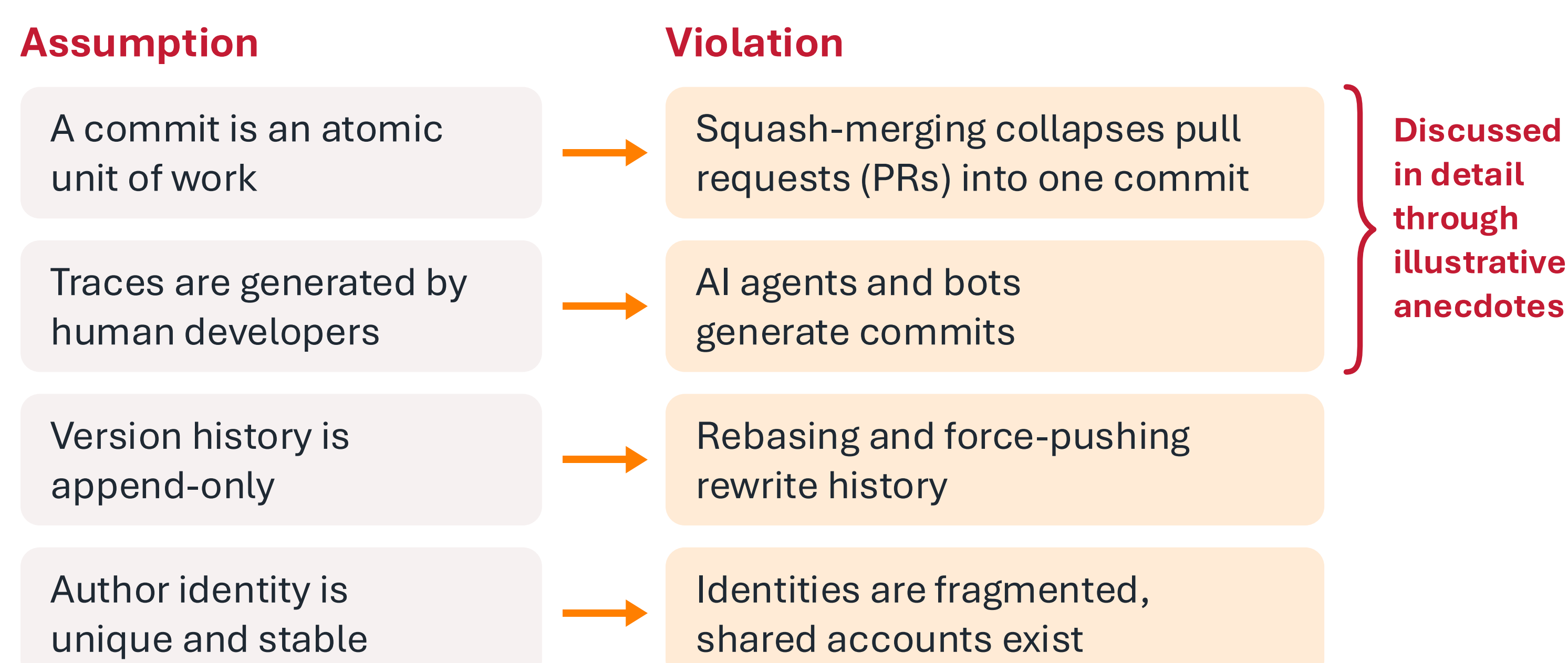
Illustrated for 8 widely-used GitHub projects that (at least, partially) use squash-merging

Project	Commits per PR (% of PRs in the project)					>1 commit	Commits per PR		# PRs
	1	2-5	6-10	11-20	>20 commits		avg.	max.	
transformers	35%	39%	12%	7%		65%	7.3	846	19,353
deno	38%	43%	11%			62%	4.3	389	15,142
TypeScript	48%	38%	8%			52%	3.8	438	14,877
flutter	51%	34%	8%			49%	3.6	1,398	48,051
bootstrap	57%	36%				43%	3.3	9,187	9,259
electron	61%	31%				39%	2.8	1,256	25,669
vscode	62%	30%				38%	2.8	3,896	47,094
react	65%	28%				35%	2.6	1,900	12,836

Time window: From project start on GitHub until May 12, 2026. Only merged PRs are considered.

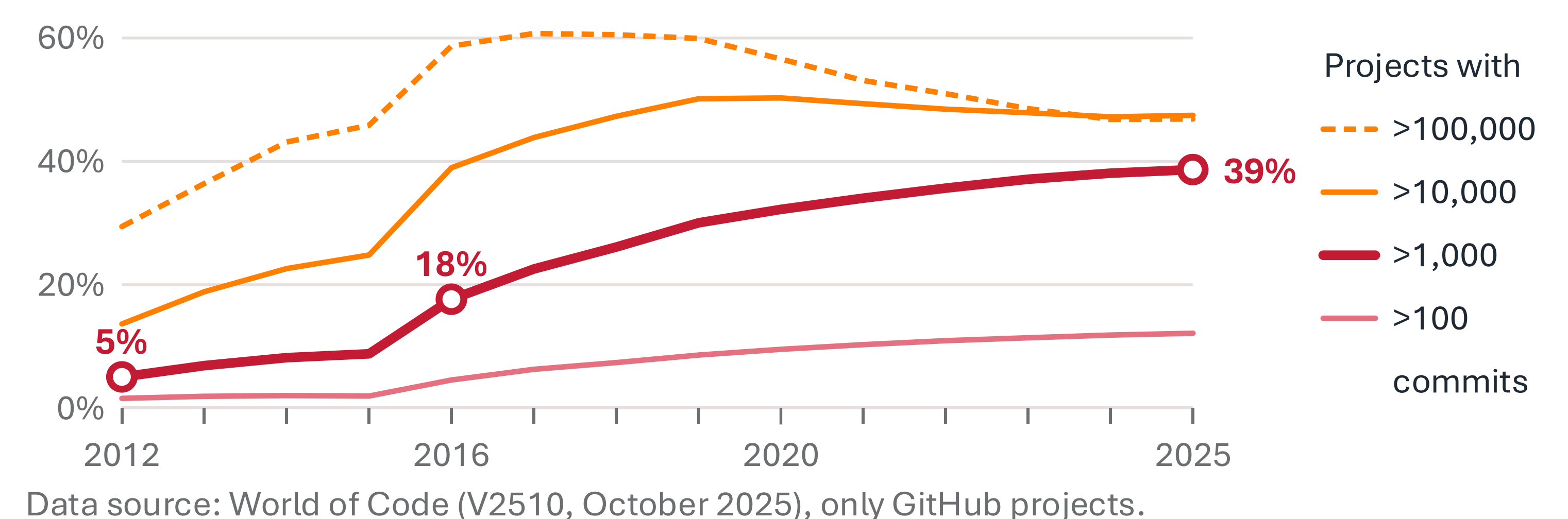
Assumptions of Data Collection Practices

Foundational assumptions of trace data are no longer valid, for example:



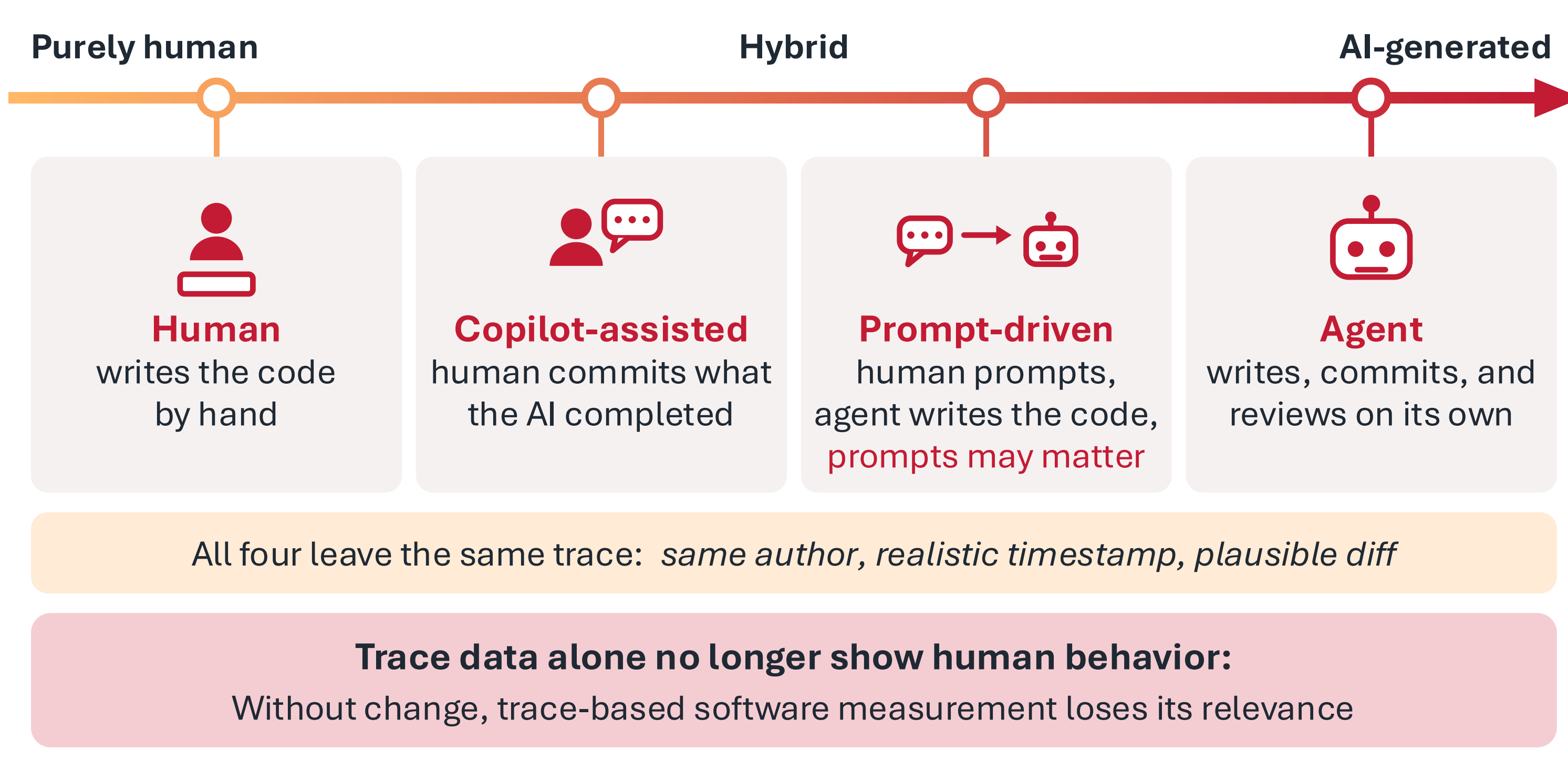
The adoption of squash-merging over time at population scale on GitHub:

Share of GitHub projects with at least one squash-merged commit



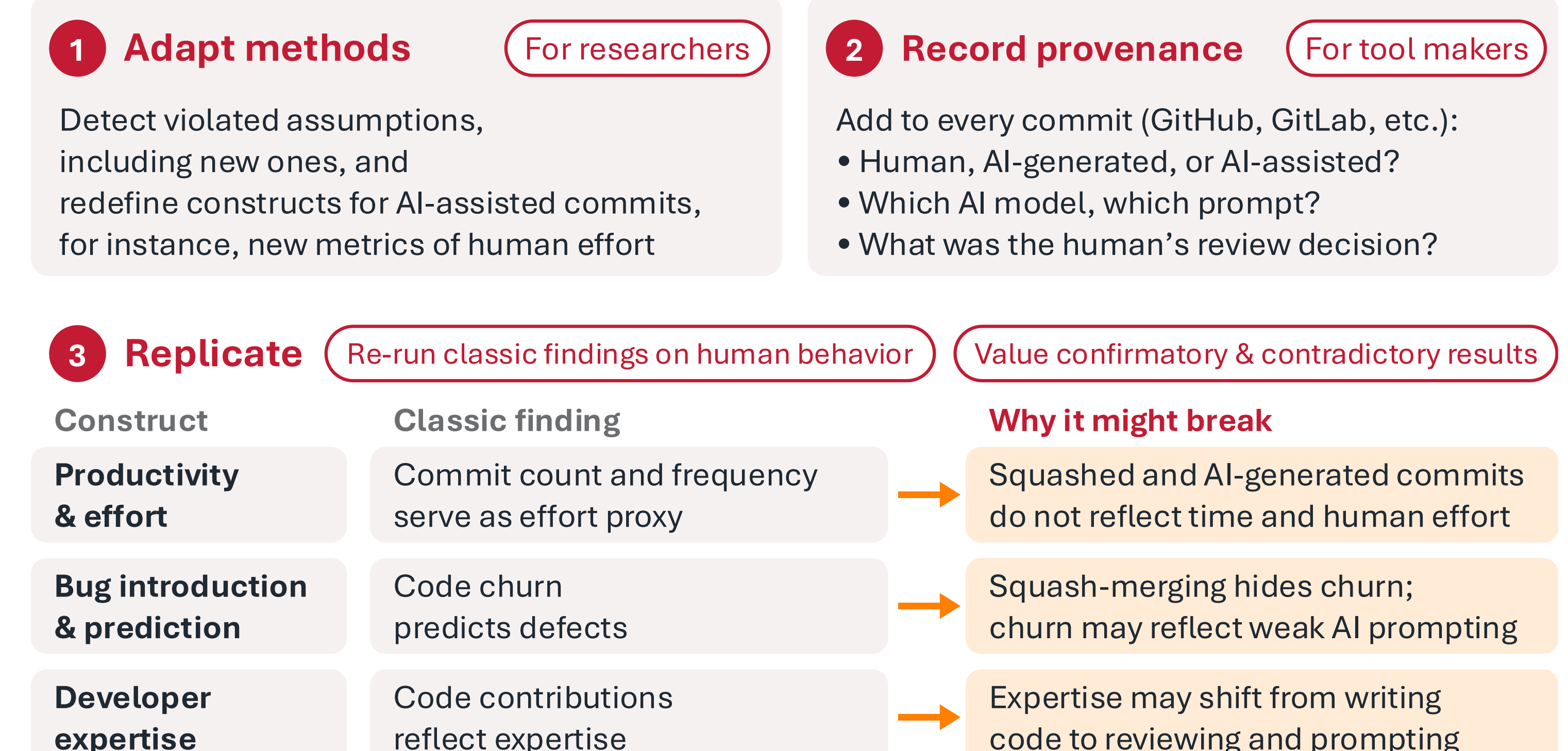
Human-Generated vs. AI-Generated Data

The distinction between “human” and “artificial” is no longer binary, but a spectrum:



Systematic AI-Assisted Replication Agenda

Which findings established under earlier conditions are replicable with today's data?



AI-assisted replication workflow



Human validation, interpretation of discrepancies, and assessment of adaptations

Call for action: Use AI-assisted tools to make replication feasible at scale

