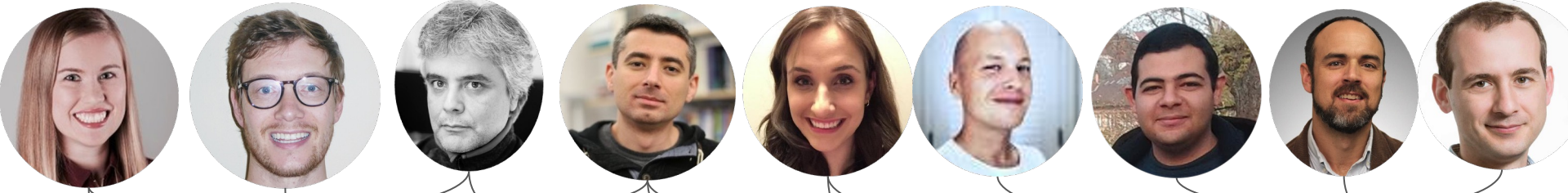
Mining Software Repositories: A Research Journey

Mining Software Repositories
Lisbon, Portugal, April 15th 2024

Margaret-Anne Storey
Foundational Contribution Research Award

@margaretstorey

University of Victoria



Screen Casts

Code Review

Stack Overflow

GitHub

Bots

Build

Twitter

CI

Thriving

Testing

DevEx

Productivity

Covid

Security

Tagging

Dashboards

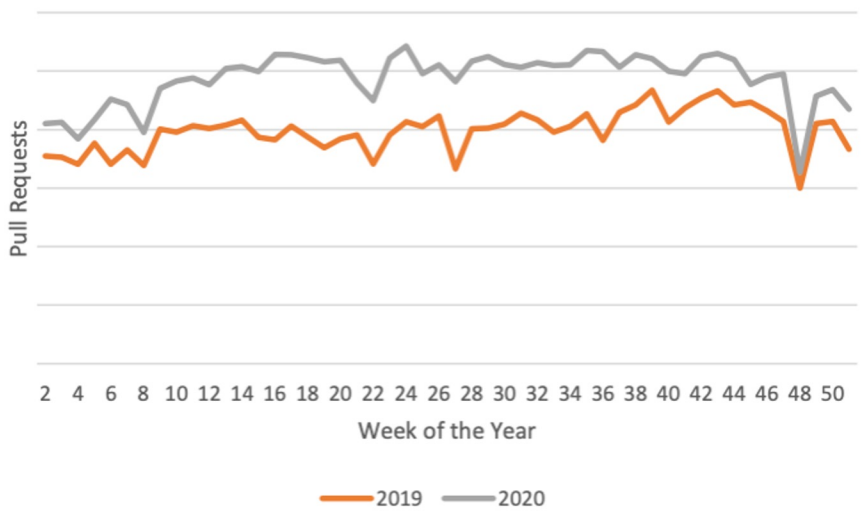


Figure 3: Weekly pull requests over time.

https://www.microsoft.com/en-us/research/uploads/prod/2021/01/NewFutureOfWorkReport.pdf p. 28

3 | Lived experiences of developers

Selecting Empirical Methods for Software Engineering Research

(Easterbrook, Singer, Storey, Damian,
Idea formed at CSCW 2006, Banff)

https://link.springer.com/book/10.1007/978-1-84800-044-5

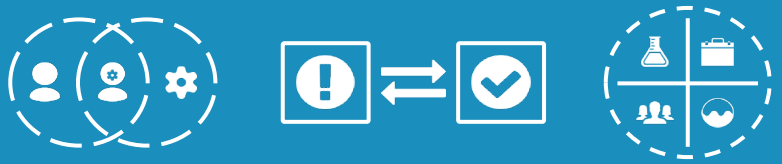
Inspired by: Extending: Runkel & McGrath: Research on Human Behavior: A Systematic Guide to Method, 1972

"A paradigm is a shared world view that represents the beliefs and values in a discipline and that guides how problems are solved."

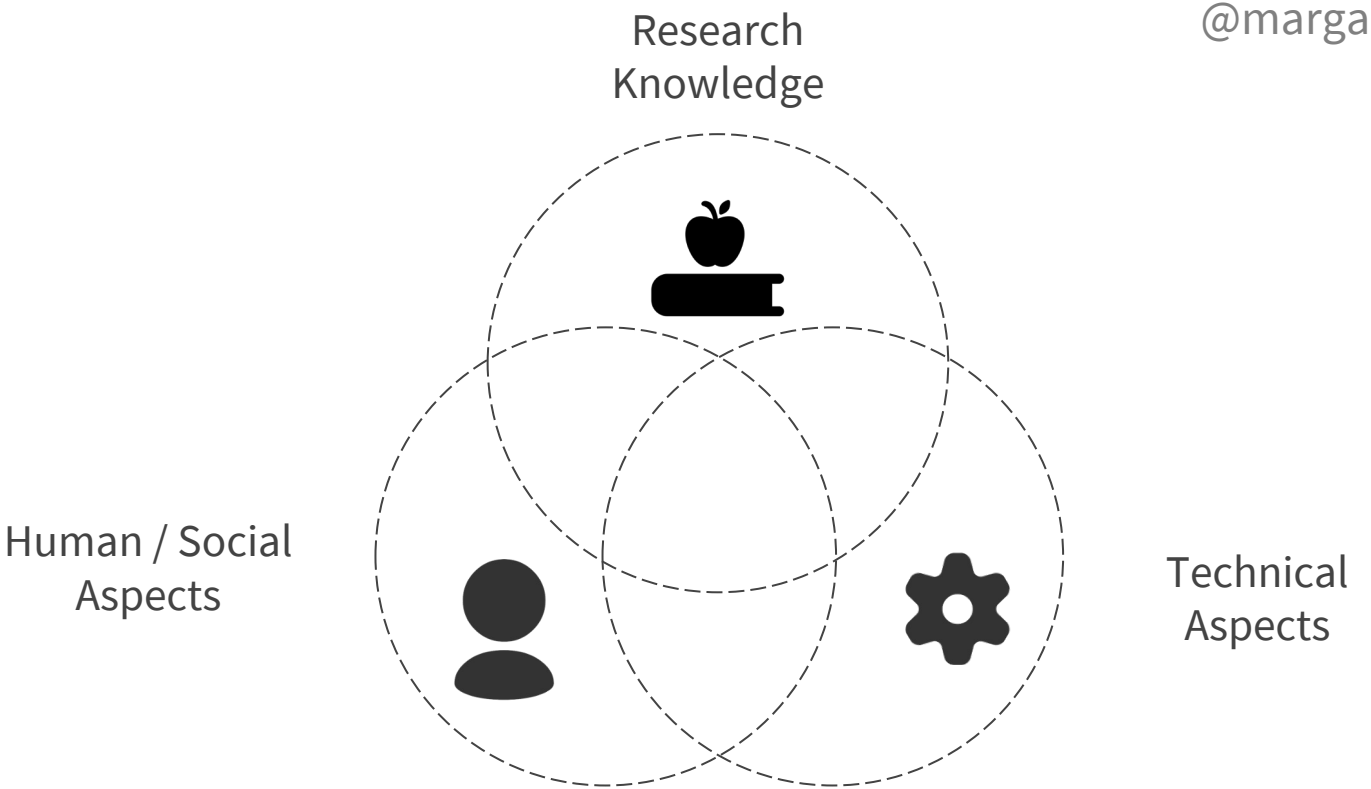
– Schwandt, 2001

https://doi.org/10.1007/s10664-020-09858-z

The who, what, how of software engineering research: a socio-technical framework



Storey, Ernst, Williams, Kalliamvakou
Empirical Software Engineering, 2020



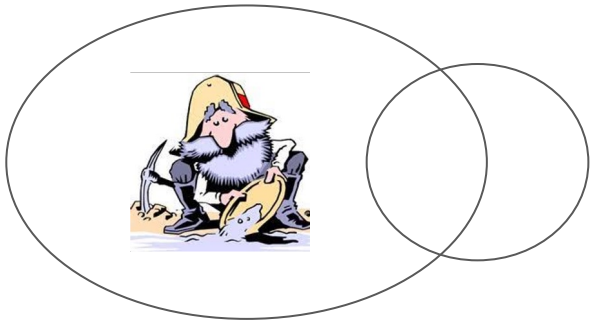
 | ***Who*** is the **beneficiary** of our research?

Descriptive (Problem)

Theories about what is or what could be

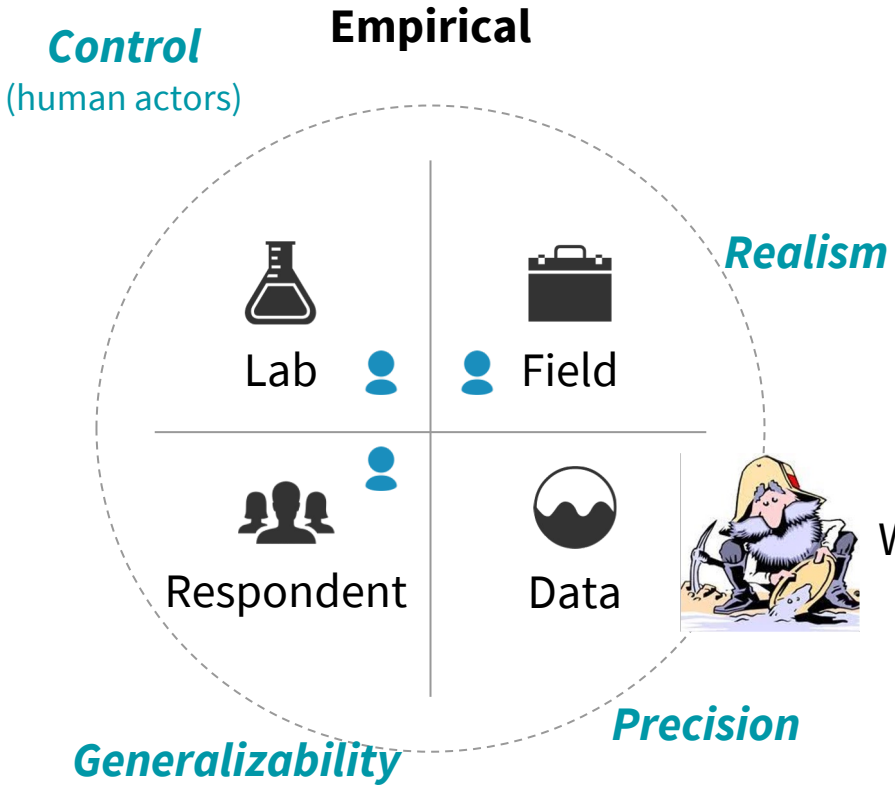
Solution

Tools or processes to improve something or fix a problem



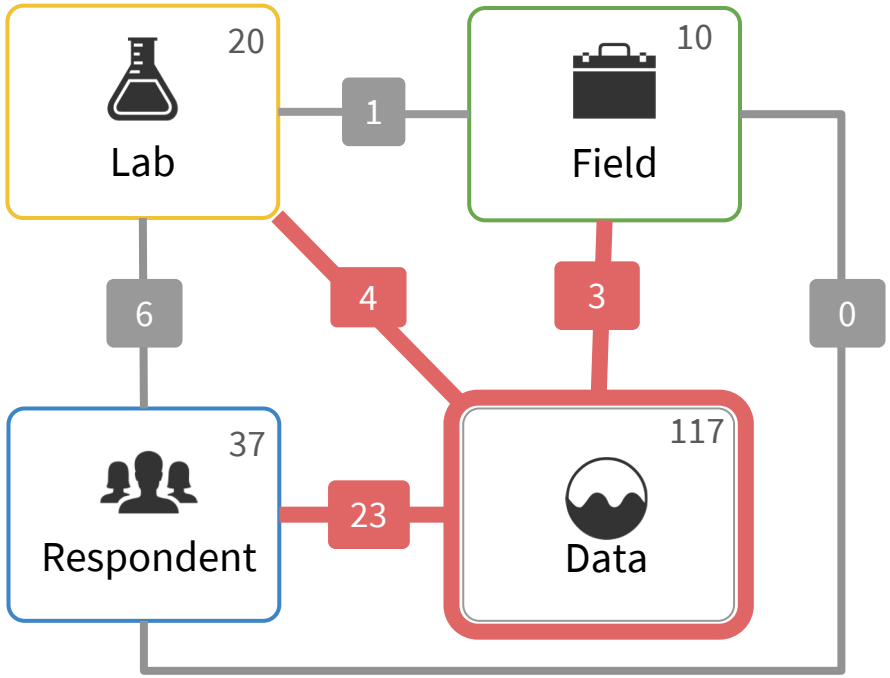
8 | ***What*** kinds of **questions** do MSR researchers answer?

Engström, Storey, Runeson, et al. How software engineering research aligns with design science, Empir Software Eng 2020



What are we mining for?

How do we answer our research question, trade-offs we make



10 **Mixed Method (Triangulation)** across 151 SE papers back in 2017!

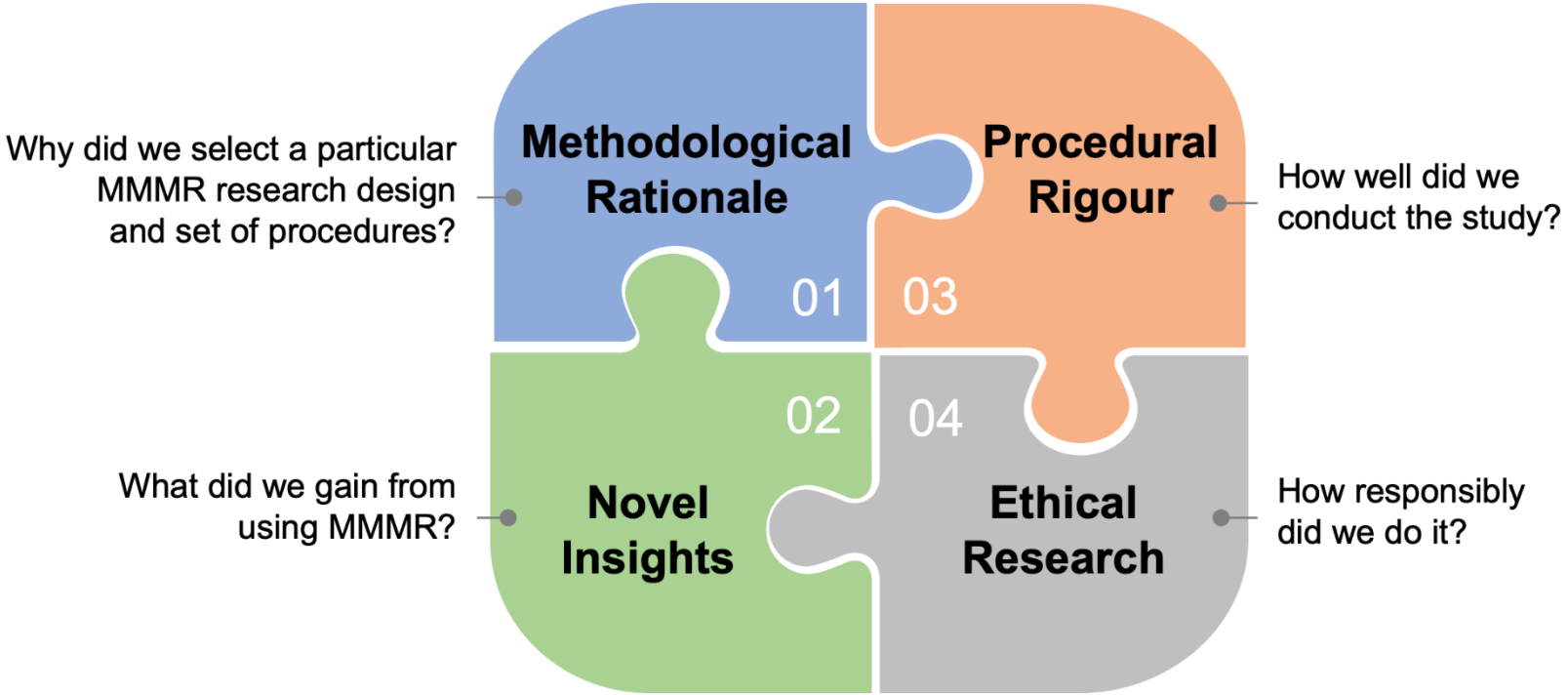
Guidelines for Using Mixed and Multi Methods Research in Software Engineering

M. Storey, R. Hoda, A. Milani, T. Baldassarre

https://bit.ly/mixedmethodsSE

A research approach where **multiple strategies** are used to collect, analyse, and integrate data to address a research problem. For **mixed** methods, **both qualitative and quantitative** data are integrated to produce novel insights.

Mixed and multi-methods research (MMMR)



Four **guiding principles**

Inspired by: Cheryl N. Poth. Innovation in mixed methods research: a practical guide to integrative thinking with complexity. SAGE, Los Angeles, 2018.

Planned or **Emergent** Design

Inductive/Deductive Dominance

Timing of Procedures

Data **Integration** (when / how)

Characteristics of mixed/multi method designs

Sequential **exploratory** strategy

Sequential **explanatory** strategy

Convergent parallel

Embedded

Multi-method

Sequential Explanatory Design:

Sam

(Initial) Hypothesis: A new code review process improves time-to-merge of code changes and improves the quality of the reviewed code.

Data mining *(deductive)*

quaN Collection → quaN Analysis

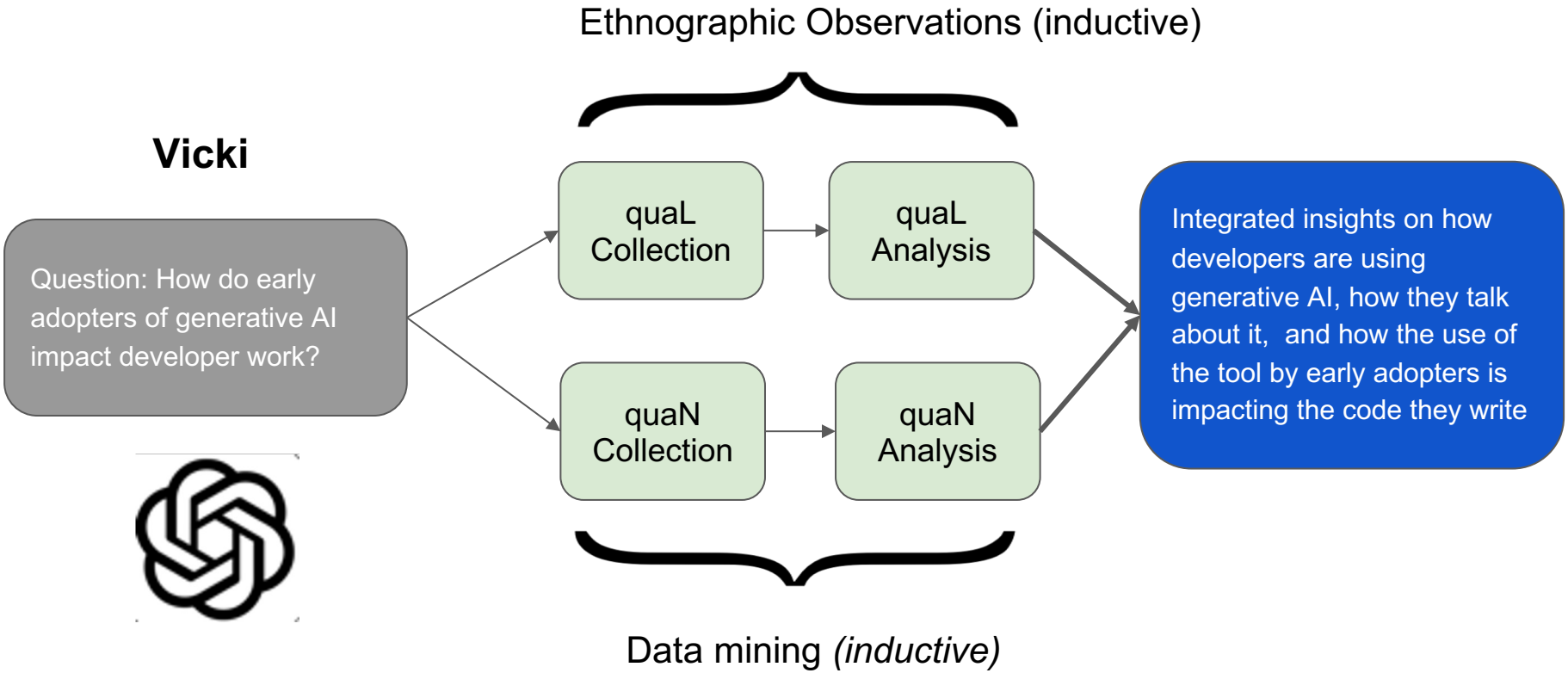
Evidence for the hypothesis is not found, leading to (unplanned) interviews to understand the negative results

Interviews *(inductive)*

quaL Collection → quaL Analysis

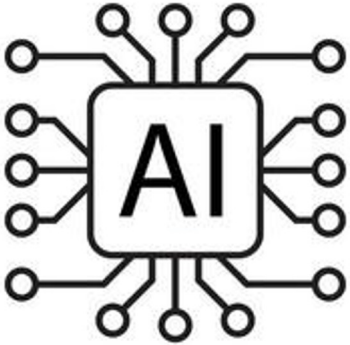
Insights about why a new code review process does not seem to improve the time-to-merge or quality of code changes

Convergent Parallel Design:

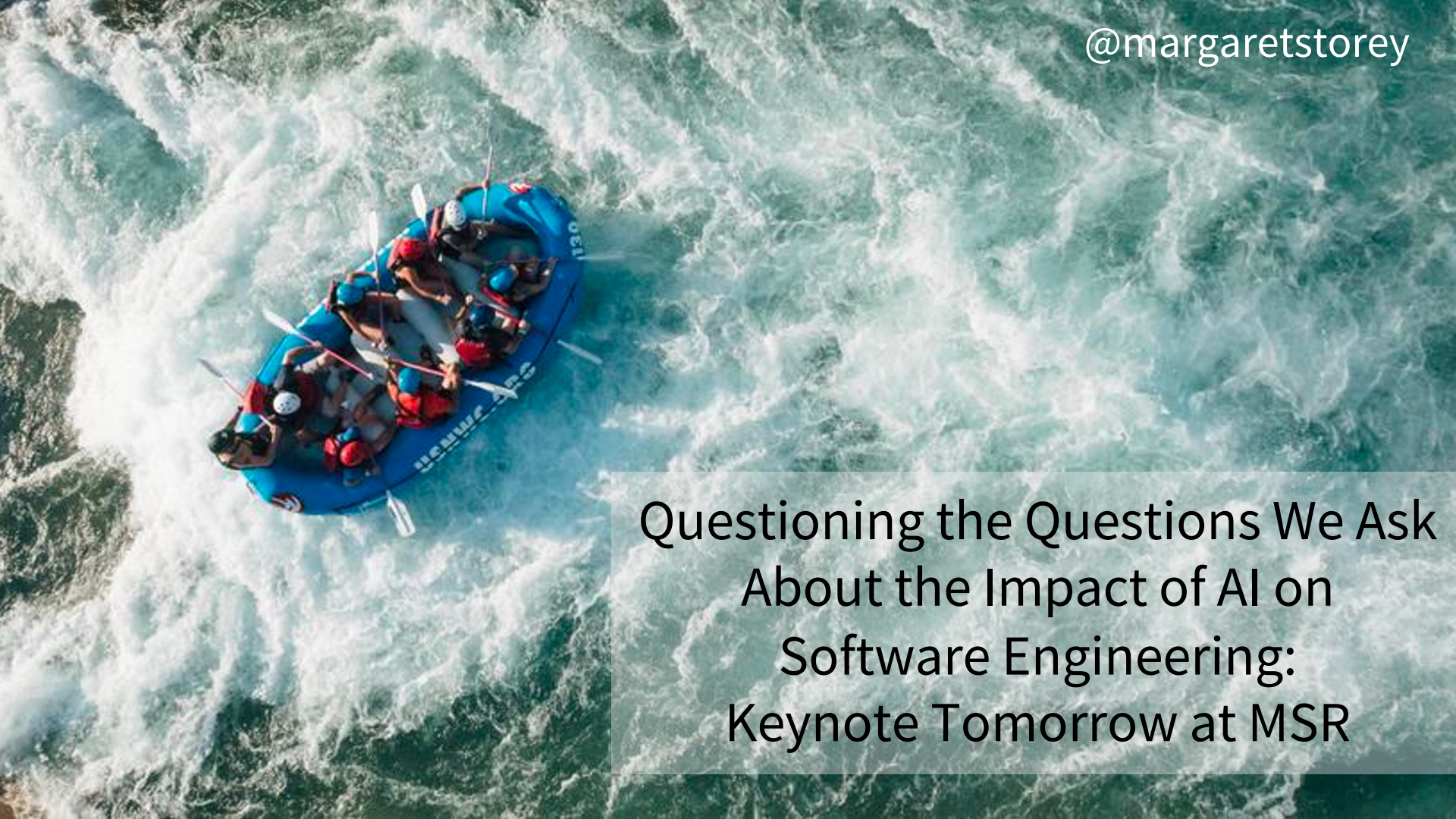


18 | Mixed Method **Study Design Anti-Patterns**

+


AI

= ?


@margaretstorey
Questioning the Questions We Ask About the Impact of AI on Software Engineering: Keynote Tomorrow at MSR