How much does AI impact development speed?

An enterprise-based randomized controlled trial

Paradis et al.
ICSE-SEIP 2025

Problem

Significant investment in AI code-editing tools, but unclear how these affect productivity.

Gap

What is the impact of AI on *time spent* on tasks in *enterprise context*?

Hook

A robust estimate of the impact of AI tools is critical to their long-term adoption.

Research Questions

RQ1: What impact does AI have on time spent completing an enterprise-grade development task?

RQ2: How do developer and task characteristics influence our estimates of the impact of AI assistance on time spent on task?

RQ3: How do developer and task characteristics interact with the use of AI to accelerate or slow down certain developers and not others?

Study Overview

Randomized Controlled Trial (RCT)

Participants were randomly assigned either to use or not use AI tools to solve a task.

Data Analysis

T-tests and linear regressions to assess AI impact on speed on task.

Hypotheses Testing

Answer RQs by testing hypotheses built from theoretical framework and literature.

Randomized Controlled Trial
Participants

96 full-time software engineers from Google that met the following criteria:

- Working at Google for >= 1 years
- Proficient in C++
- Submitted code to Google's repo in the past
- Used Cider V as their main IDE
- Some experience with task domain

Randomized Controlled Trial
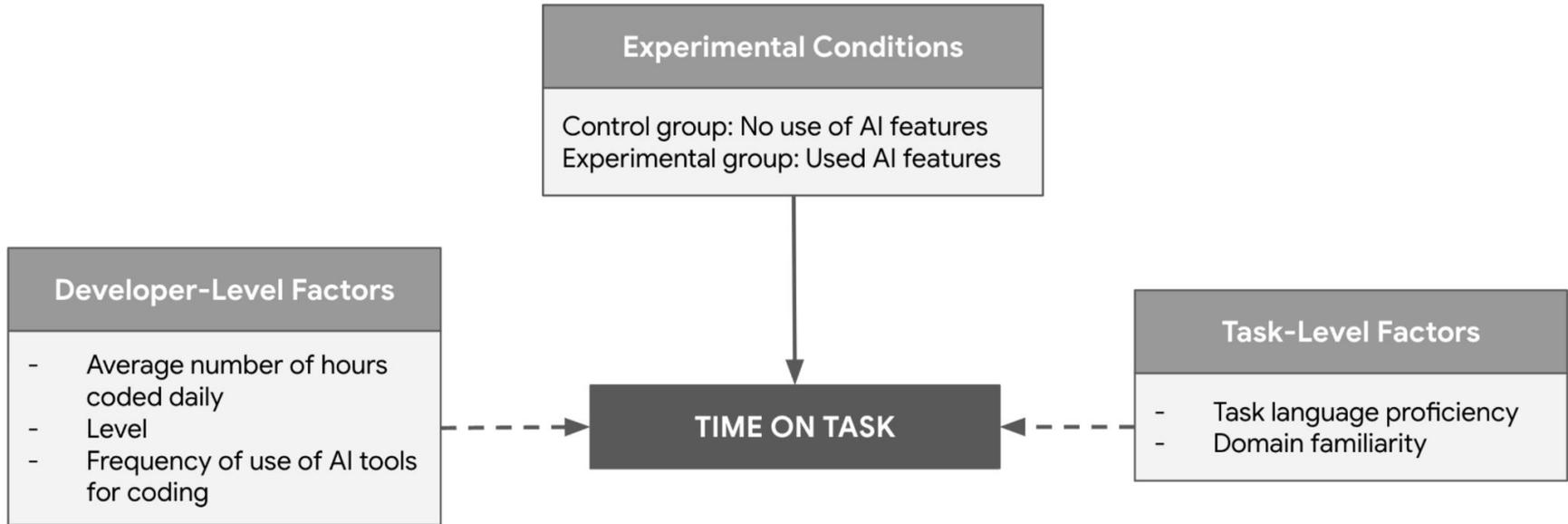## Tasks
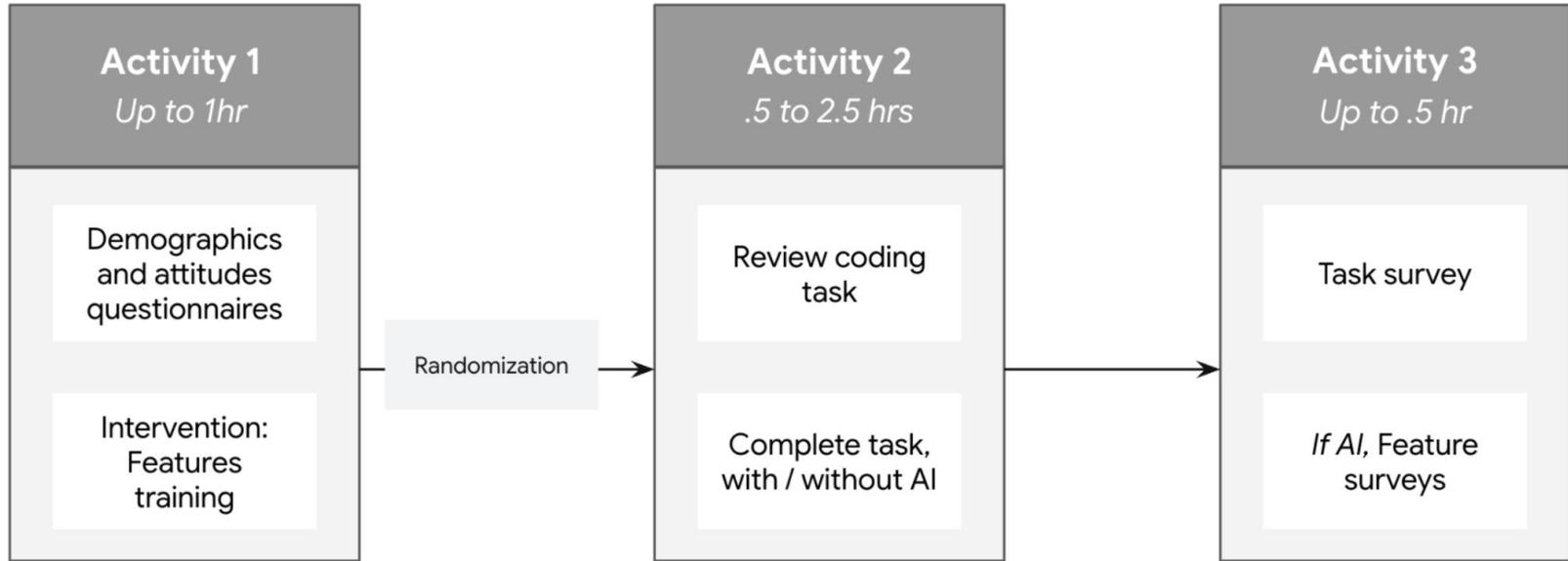
Randomized Controlled Trial
Choosing a task

Task: implementing a fairly complex logging service

- Reflects real-world tasks
- Leverages the full range of developer tools
- Feasible within the context of the study

Task was pre-tested with eight engineers.

Hypotheses Building
Theoretical Framework

Hypotheses Building
Variables

Dependent variable: Time spent on task

Independent Variable: Use of AI (ExpCon)

Confounding variables

- Developer-level
  - Average number of daily hours on coding (AvgProgHrsDay)
  - Seniority (Level)
  - Frequency of AI use (NbrHighFreAreas)
- Task-level
  - Proficiency with task lang (TopLangCpp)
  - Task domain expertise (DataLogExp)

Hypotheses

For RQ1:
(H1) Participants assigned to the AI condition will spend less time on task than those assigned to the control group.

For RQ2:
(H2) Controlling for confounding developer-level and task-level variables, participants assigned to the AI condition will still spend less time on task than those assigned to the control group.

Hypotheses

For RQ3:

(H3) There will be a negative interaction effect between the experimental condition and average daily hours spent coding.

(H4) There will be a negative interaction effect between the experimental condition and seniority.

(H5) There will be a negative interaction effect between the experimental condition and the frequency with which developers use AI coding tools.

Analytic Approach

Two-tailed Student t-test on logged time on task, comparing mean time for participants in the control versus experimental groups.

To assess robustness of the estimate, they used three linear regressions:
- Experimental cond. + developer-level + task-level factors
- Experimental cond. + developer-level factors
- Experimental cond. + task-level factors

Analytic Approach

To explore interaction between experimental cond. and developer-level factors: separate linear regressions with interaction effects, one for each covariate.
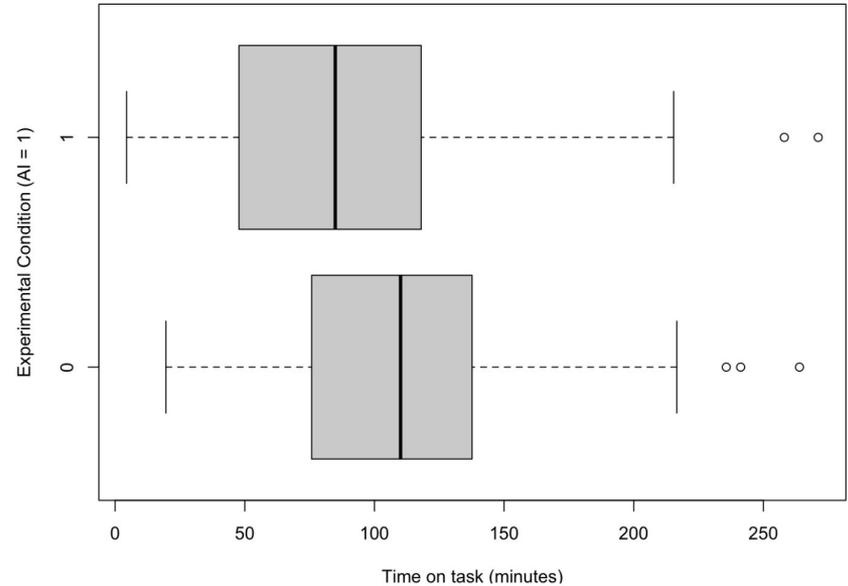
- Average hours coded daily
- Seniority
- Frequency of use of AI tooling

Results

RQ1: Impact of AI tools on developer task speed

Hypothesis 1: Supported

Using AI was associated with a shorter time on task ($p < 0.05$).

RQ2: Impact of covariates on main effect

Hypothesis 2: Partially Supported

Controlling for other factors, using AI remained associated with a shorter time on task but lost its significance ($p = \text{NS}$).

RQ3: Interaction effects between experimental condition and other factors

Hypothesis 3: Partially supported

Interaction effect was large and negative but not statistically significant, and model fit for Model 4 was significant (Adjusted $R^2 = 0.075; p = 0.018$).

Hypothesis 4 and 5: Rejected

Interaction effects between the use of AI features and the seniority or the frequency with which developers use AI coding tools were both negative but not statistically significant.

Takeaways
Main effect

- Developers who used AI were statistically significantly faster than those who did not.
- Similar effect size when controlling for other factors, but not statistically significant (p = 0.086).
- Estimate of 21% increase in development speed attributable to AI.

Takeaways
Interaction effects

- Developers who code more per day may be faster with AI that those who code less -> points to research to make more junior developers even faster with AI.
- No statistically significant evidence that more familiarity with AI tools or seniority increases speed on task.

Limitations

- A "true estimate" of AI impact on developer speed is a moving target.
- Larger sample sizes are needed for more confident results on interaction effects between use of AI tools and developer-level factors.
- Risk to external validity: participants are all Google employees.
- Inherent mismatch between experiment task and real-world work.

How much does AI impact development speed?

An enterprise-based randomized controlled trial

Paradis et al.
ICSE-SEIP 2025