Generative AI at Work

Lindsey Raymond (MIT)

Joint with
Erik Brynjolfsson (Stanford & NBER)
and
Danielle Li (MIT & NBER)

Why care about machine learning?

Traditionally: Computers require instructions

Why care about machine learning?

Traditionally: Computers require instructions

Why care about machine learning?

Traditionally: Computers require instructions

- Use inputs X and instructions $f(X)$ to produce output Y

Impacts of IT shaped by routine nature of work

- Some tasks follow well-understood procedures
  - e.g. numerical processing, executing a sequence of instructions

Why care about machine learning?

Traditionally: Computers require instructions

- Use inputs X and instructions $f(X)$ to produce output Y

Impacts of IT shaped by routine nature of work

- Some tasks follow well-understood procedures
  - e.g. numerical processing, executing a sequence of instructions
- Wage and employment effects felt in areas like information processing and clerical work

Why care about machine learning?

ML learns instructions:

Why care about machine learning?

ML learns instructions:

Why care about machine learning?

ML learns instructions:

- Requires specification of a goal

Why care about machine learning?

ML learns instructions:

- Requires specification of a goal
- Uses examples–inputs and outputs (X, Y)–to infer instructions $f(X)$

Why care about machine learning?

ML learns instructions:

- Requires specification of a goal
- Uses examples–inputs and outputs (X, Y)–to infer instructions $f(X)$

Knowledge is an output, not an input

Why care about machine learning?

ML learns instructions:

- Requires specification of a goal
- Uses examples–inputs and outputs (X, Y)–to infer instructions $f(X)$

Knowledge is an output, not an input

- Can use ML to recognize faces, diagnose heart attacks, find high-potential employees

Why care about machine learning?

ML learns instructions:

- Requires specification of a goal
- Uses examples–inputs and outputs (X, Y)–to infer instructions $f(X)$

Knowledge is an output, not an input

- Can use ML to recognize faces, diagnose heart attacks, find high-potential employees
- All tasks without well-understood recipes

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

- Growing literature suggesting AI can outperform humans on average

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

- Growing literature suggesting AI can outperform humans on average

Learn what makes people good at their job

- Most training data generated by people doing their jobs

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

- Growing literature suggesting AI can outperform humans on average

Learn what makes people good at their job

- Most training data generated by people doing their jobs
- Large and persistent performance differences among people

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

- Growing literature suggesting AI can outperform humans on average

Learn what makes people good at their job

- Most training data generated by people doing their jobs
- Large and persistent performance differences among people
- ML might be able to learn which behaviors are associated with success
  - Including **tacit** rules top performers apply but may not explicitly understand

Generative AI: predict how a good worker would **behave**

What's possible when machines don't need instructions?

Expands the set of tasks machines can perform

- Growing literature suggesting AI can outperform humans on average

Learn what makes people good at their job

- Most training data generated by people doing their jobs
- Large and persistent performance differences among people
- ML might be able to learn which behaviors are associated with success
  - Including **tacit** rules top performers apply but may not explicitly understand

Generative AI: predict how a good worker would **behave**

- We can do more than classify, we might be able to generate what a good worker would say or do

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?
- Generative AI is coherent but inaccurate

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?
- Generative AI is coherent but inaccurate
- How will it fit into organizations that rely on humans?

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?
- Generative AI is coherent but inaccurate
- How will it fit into organizations that rely on humans?

This paper: the first large-scale study of generative AI in the workplace

- Can access to generative AI assistance increase worker productivity?

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?
- Generative AI is coherent but inaccurate
- How will it fit into organizations that rely on humans?

This paper: the first large-scale study of generative AI in the workplace

- Can access to generative AI assistance increase worker productivity?
- For which workers and why?

How will generative AI shape the workplace?

Little evidence and lots of reasons to be skeptical

- Lots of exciting tech in the lab that hasn't had big economic impacts
- Are distinctions between top and bottom workers machine-learnable?
- Generative AI is coherent but inaccurate
- How will it fit into organizations that rely on humans?

This paper: the first large-scale study of generative AI in the workplace

- Can access to generative AI assistance increase worker productivity?
- For which workers and why?
- How does adoption change the experience of work?

Generative AI in the wild

Setting: Technical customer support chat

- One of the top use cases for modern AI tools

Technology: Conversational customer support assistant

- Provides real-time recommendations for how to communicate

Empirical Design: Staggered roll-out in technical support for a large Fortune 500 software firm

- 3,000,000 conversations from 3,000 agents

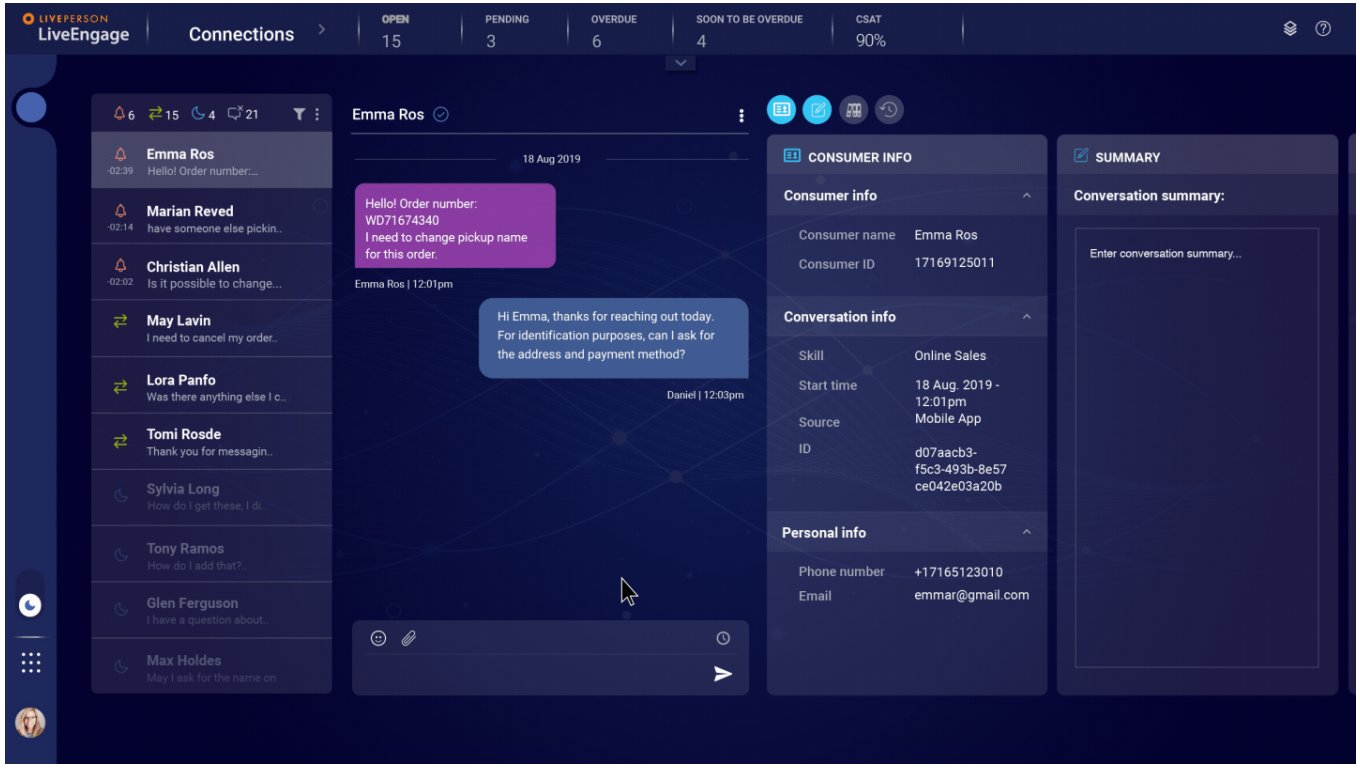
Findings:

1. Access to AI recommendations increase productivity by 14 percent
2. 35 percent for least skilled/experienced, no impact on top of the distribution
3. Suggestive evidence of AI-driven knowledge transfer

Roadmap

1. Setting and Data
   - A. Technical Support
   - B. Generative AI in Tech Support
   - C. Data and Study Design
2. Results
   - A. Productivity
     - A.1 Average effects
     - A.2 By skill and tenure
     - A.3 Learning
     - A.4 Knowledge diffusion
   - B. Experience of work
     - B.1 Customer tone
     - B.2 Attrition

Why technical support?



Technical feasibility:

- Millions of chats with automatically labeled outcomes

Business need:

- High turnover
- Knowledge-intensive, large and persistent performance differences

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as “I can’t login”?

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?
3. Recognize what gets people yelled at
   - How to communicate in ways that alleviate customer frustration?

A mix of technical, business and accounting knowledge:

- "I can't login!"

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?
3. Recognize what gets people yelled at
   - How to communicate in ways that alleviate customer frustration?

A mix of technical, business and accounting knowledge:

- "I can't login!"
- "How do I retroactively give an employee a bonus?"

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?
3. Recognize what gets people yelled at
   - How to communicate in ways that alleviate customer frustration?

A mix of technical, business and accounting knowledge:

- "I can't login!"
- "How do I retroactively give an employee a bonus?"
- "How can I mail W2's to my staff?"

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?
3. Recognize what gets people yelled at
   - How to communicate in ways that alleviate customer frustration?

A mix of technical, business and accounting knowledge:

- "I can't login!"
- "How do I retroactively give an employee a bonus?"
- "How can I mail W2's to my staff?"

Making customers "feel" like you've solved their problem:

- "My company's payroll is due in 3 hours and I cannot create timesheets... I'm so upset, this is horrible!"

Why is customer support an ML problem?

1. Diagnose the problem
   - What underlying problems present as "I can't login"?
2. Predict what will resolve it
   - What types of solutions have typically worked for this problem?
3. Recognize what gets people yelled at
   - How to communicate in ways that alleviate customer frustration?

A mix of technical, business and accounting knowledge:

- "I can't login!"
- "How do I retroactively give an employee a bonus?"
- "How can I mail W2's to my staff?"

Making customers "feel" like you've solved their problem:

- "My company's payroll is due in 3 hours and I cannot create timesheets... I'm so upset, this is horrible!"

AI system design

Model training:

1. Start from a large language model from the GPT family

AI system design

Model training:

1. Start from a large language model from the GPT family
2. Train on millions of customer-agent conversations labeled with outcomes

AI system design

Model training:

1. Start from a large language model from the GPT family
2. Train on millions of customer-agent conversations labeled with outcomes
3. Train to maximize share resolved, customer satisfaction and call duration

AI system design

Model training:

1. Start from a large language model from the GPT family
2. Train on millions of customer-agent conversations labeled with outcomes
3. Train to maximize share resolved, customer satisfaction and call duration
   - Further fine-tuned to prefer empathetic responses over just relaying technical information

AI system design

Model training:

1. Start from a large language model from the GPT family
2. Train on millions of customer-agent conversations labeled with outcomes
3. Train to maximize share resolved, customer satisfaction and call duration
   - Further fine-tuned to prefer empathetic responses over just relaying technical information

System function:

- Real-time text suggestions and links to technical material

AI system design

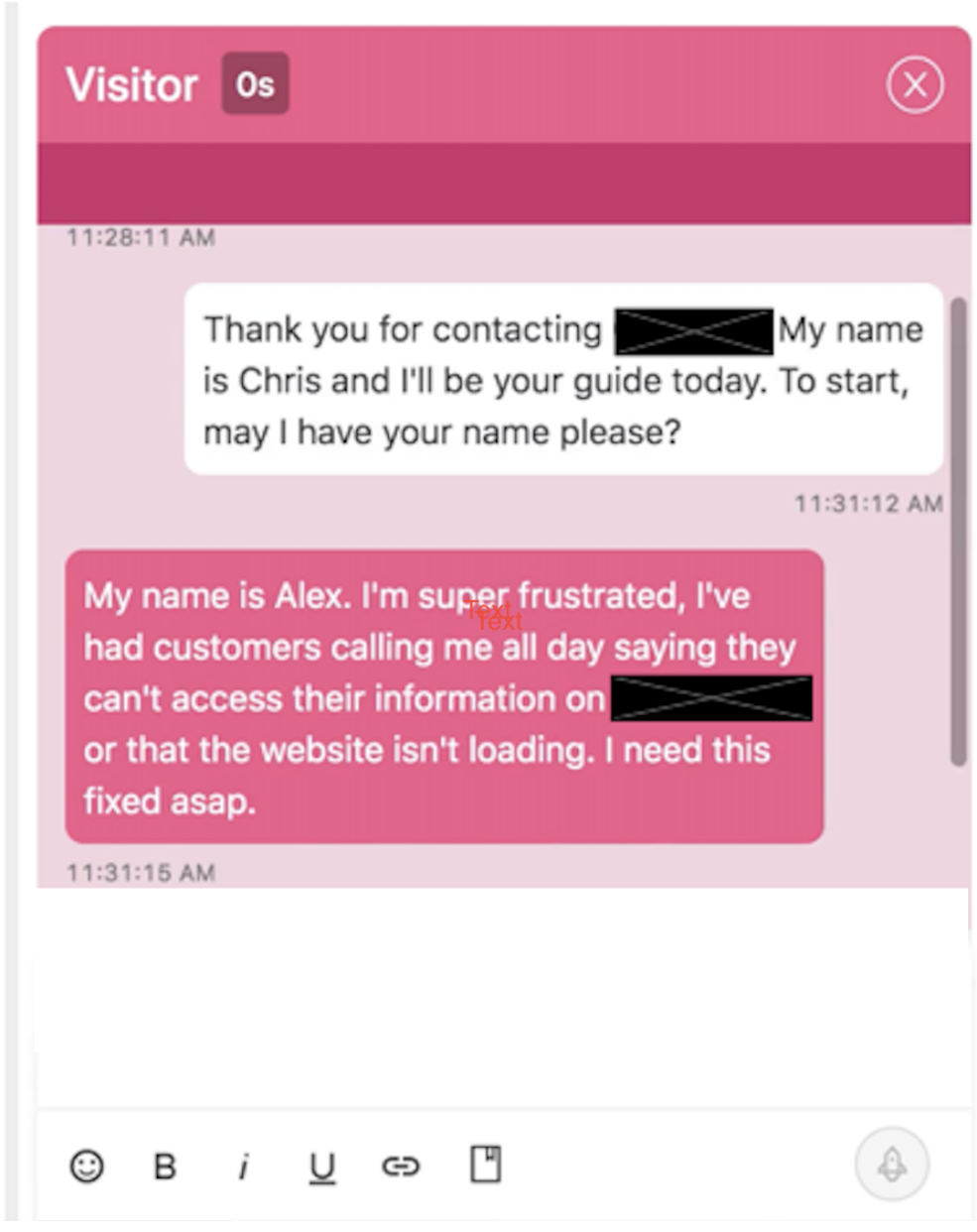
Model training:

1. Start from a large language model from the GPT family
2. Train on millions of customer-agent conversations labeled with outcomes
3. Train to maximize share resolved, customer satisfaction and call duration
   - Further fine-tuned to prefer empathetic responses over just relaying technical information

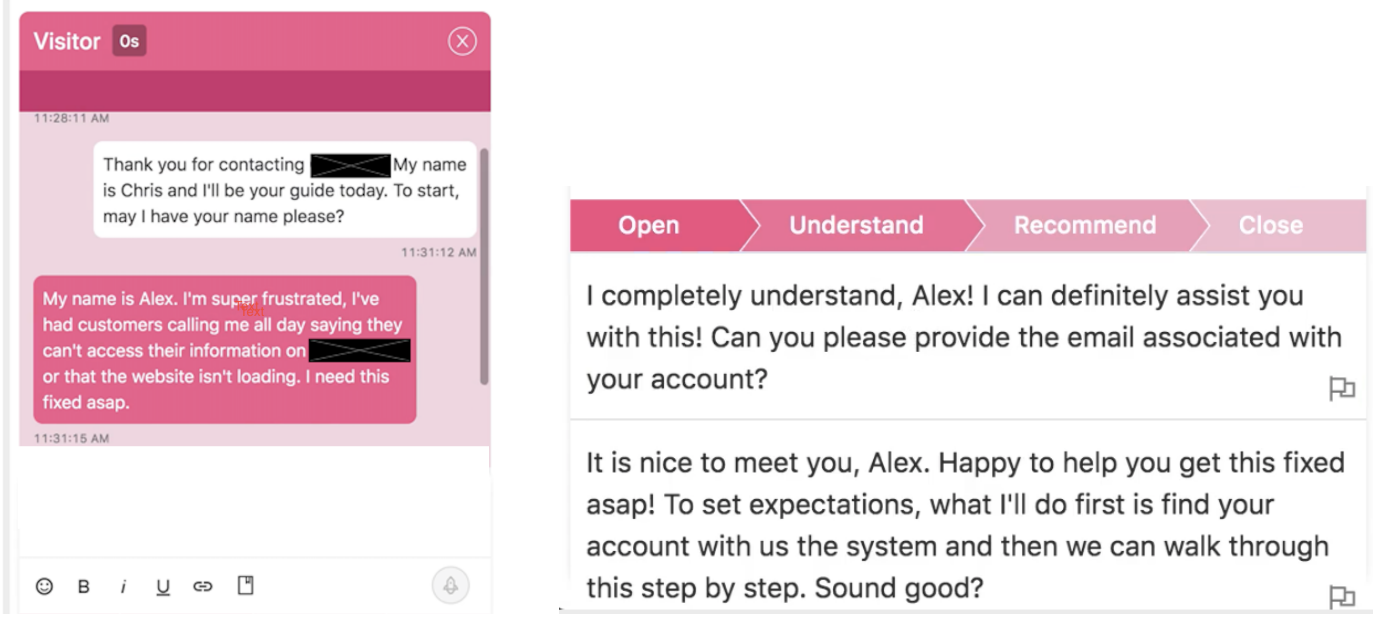
System function:

- Real-time text suggestions and links to technical material
- Agent has full discretion; customer does not see recommendations

Our tool

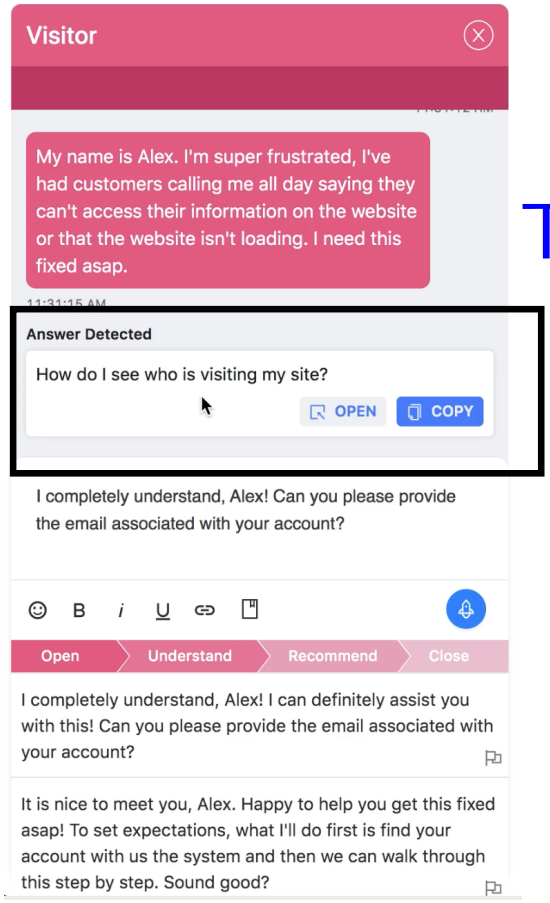

Visitor
0s
11:28:11 AM
Thank you for contacting [redacted] My name is Chris and I'll be your guide today. To start, may I have your name please?
11:31:12 AM
My name is Alex. I'm super frustrated, I've had customers calling me all day saying they can't access their information on [redacted] or that the website isn't loading. I need this fixed asap.
11:31:15 AM

AI tool provides text suggestions that the agent can use or ignore



- Recommendations based on responses that are most correlated with successful outcomes
- In this case: establishing a friendly, reassuring rapport.

Technical documentation



Technical Documentation

- Recommendations based on responses that are most correlated with successful outcomes
- In this case: links to technical documentation

Data and study design

Data

- Conversations: 3,000,000 chats between January 2020 and June 2021
- Agents: 3,000 agents, 140 teams and 5 firms
- Data: Chat text, AI output, agent interactions
- Outcomes: Issues solved per hour and customer satisfaction

Summary stats

Study Design

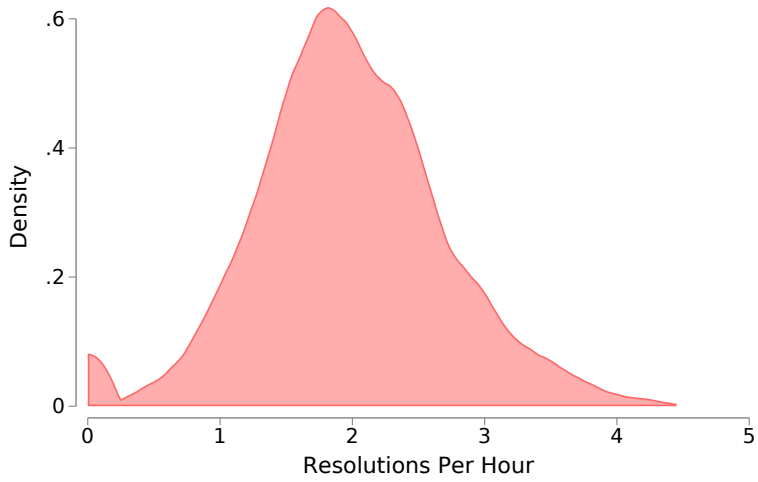
$$y_{it} = \delta_t + \alpha_i + \sum \beta_t AI_{it} + \gamma X_{it} + \epsilon_{it} \quad (1)$$

- AI rolled out over six months at the agent level Timeline
- AI_{it} indicates agent i has access to AI assistance at time t
- Use estimators robust to differential timing

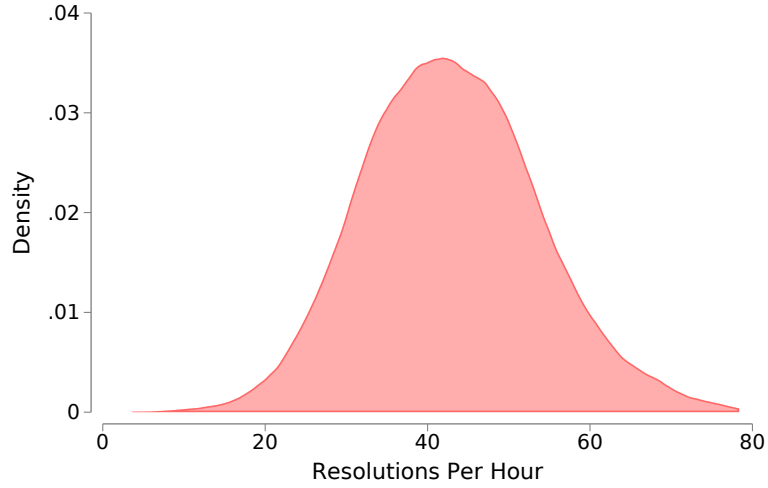
Balance table

Gains are evident in the raw data

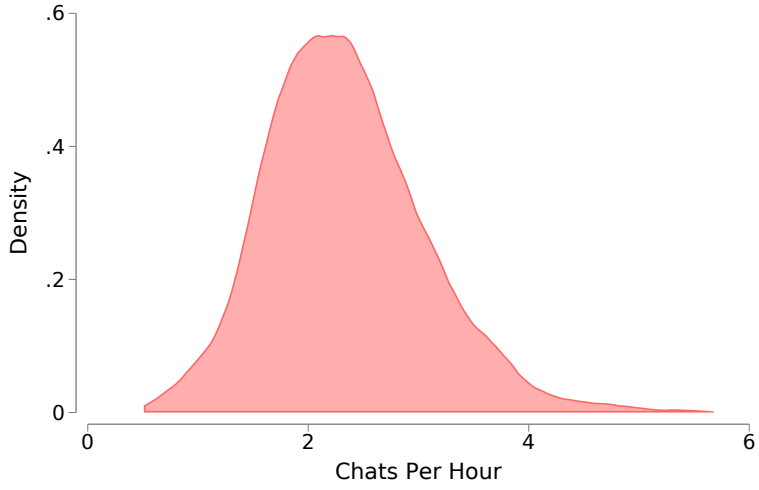
a. Resolutions per Hour



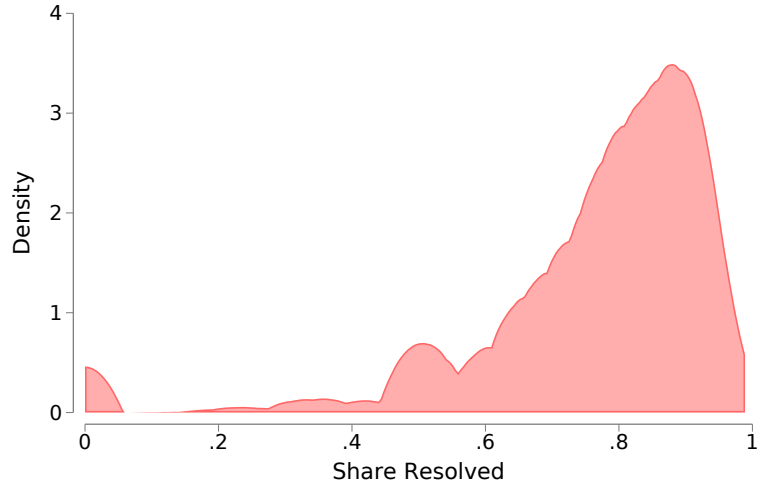
b. Average Handle Time



c. Chats per Hour

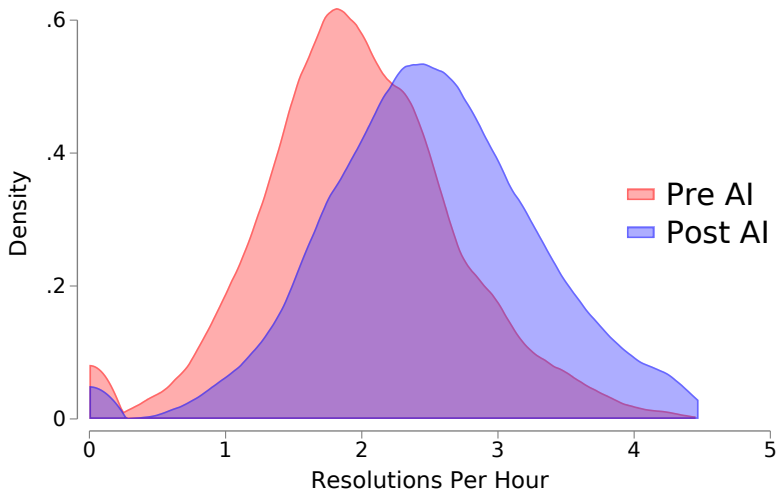


d. Share Resolved

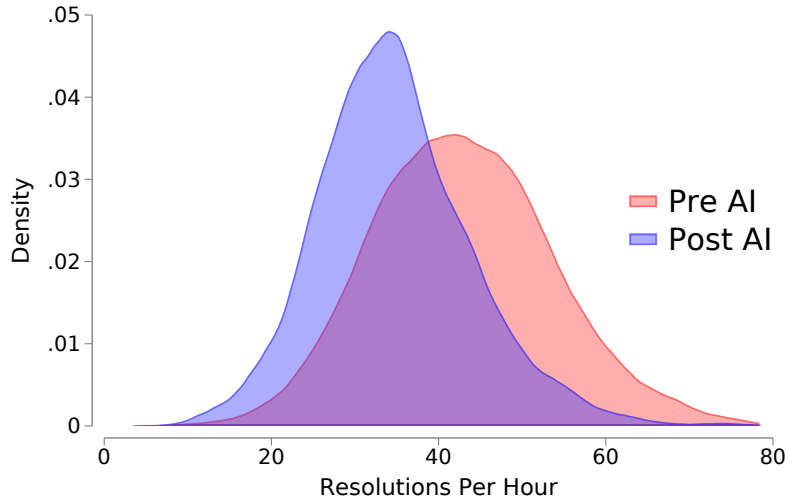


Gains are evident in the raw data

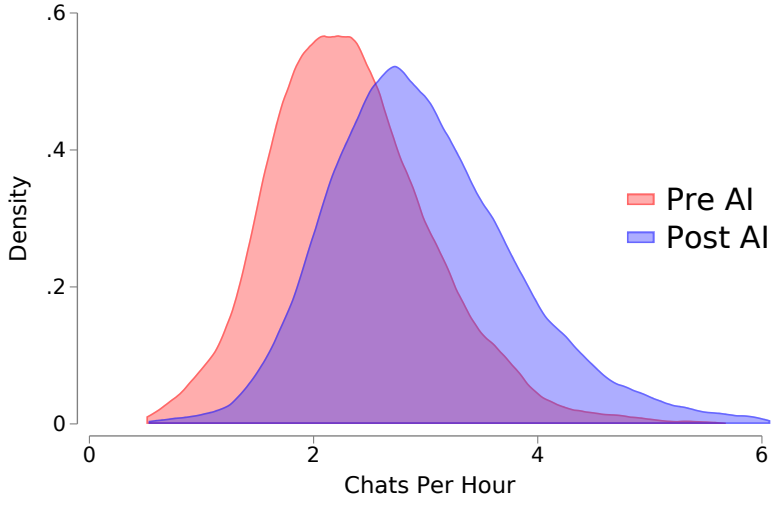
a. Resolutions per Hour



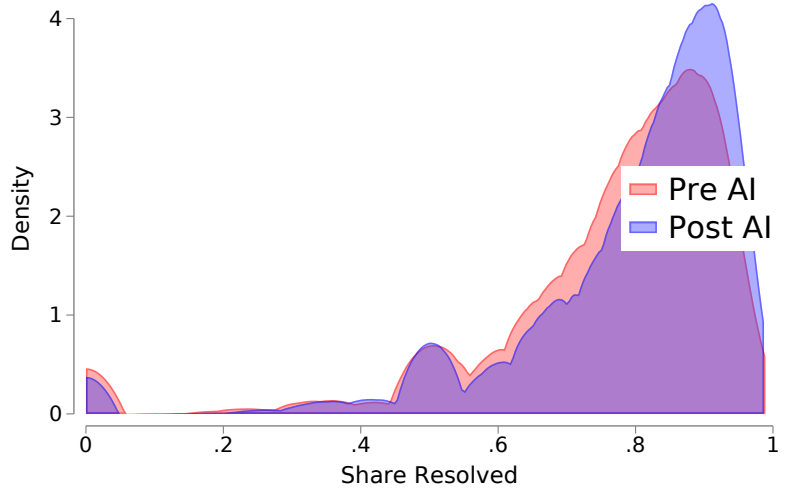
b. Average Handle Time



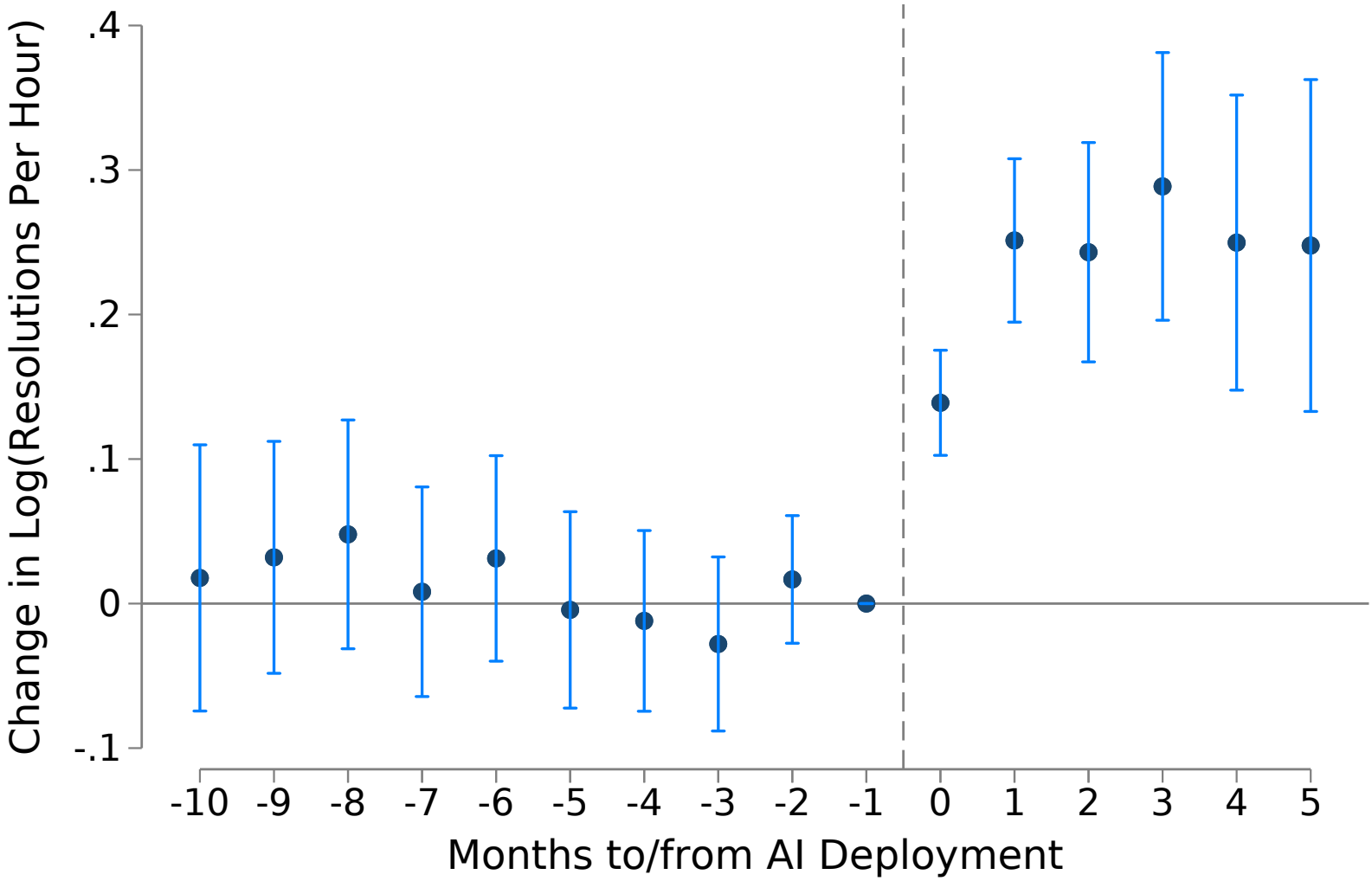
c. Chats per Hour



d. Share Resolved

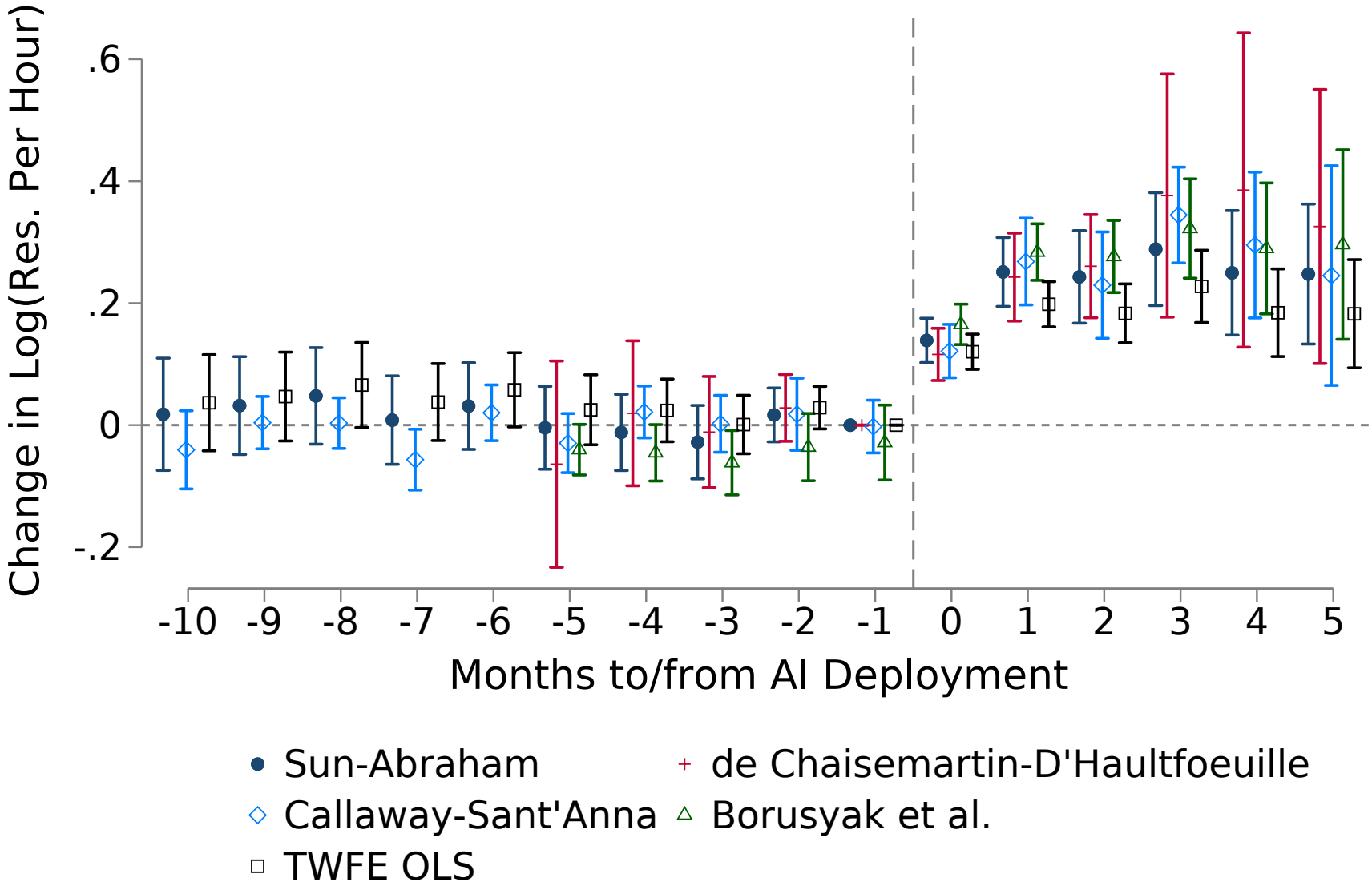


Productivity improvements persist



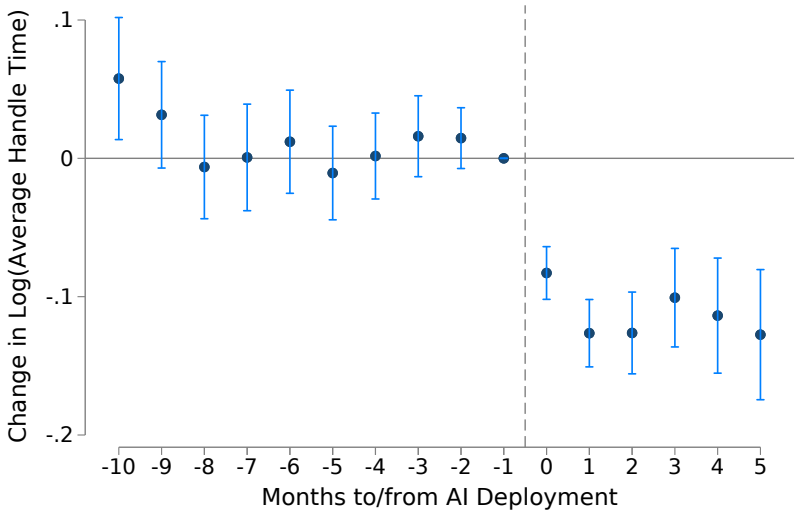
Difference-in-differences table AI adherence

Alternative estimators

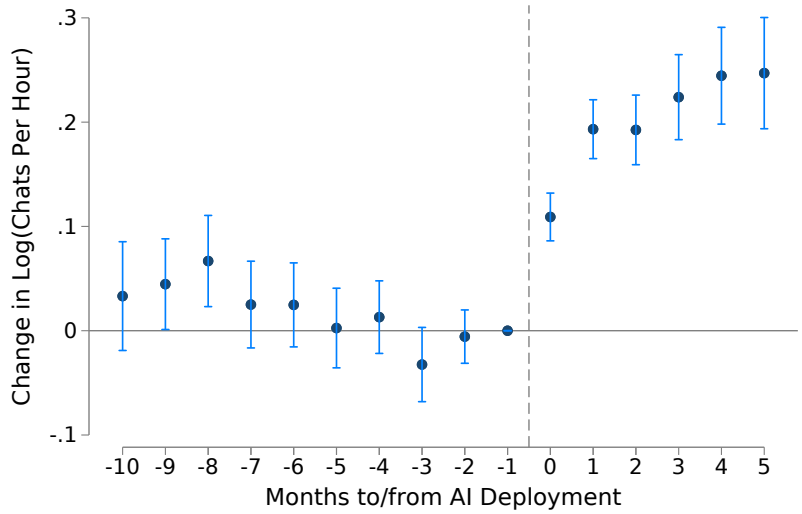


Average effects most apparent in efficiency outcomes

a. Average Handle Time

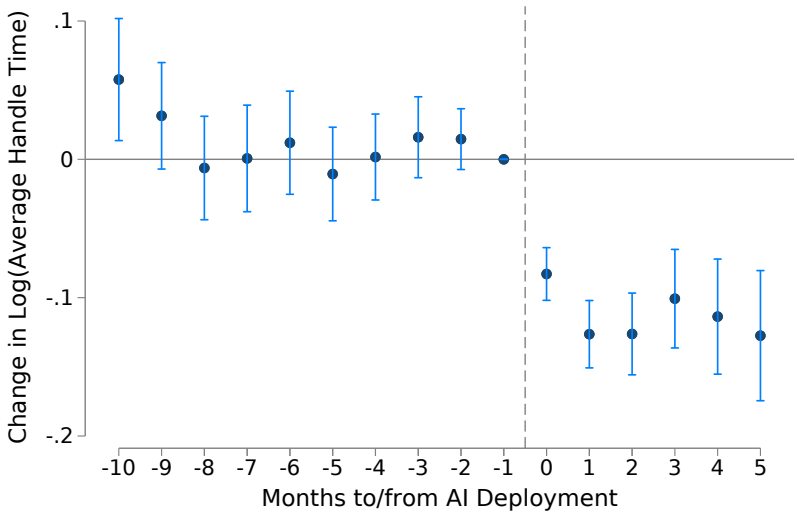


b. Calls per Hour

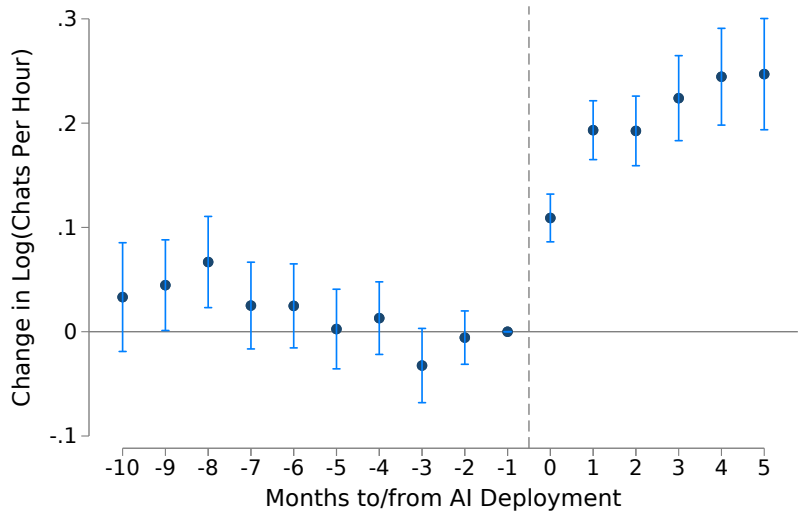


Average effects most apparent in efficiency outcomes

a. Average Handle Time

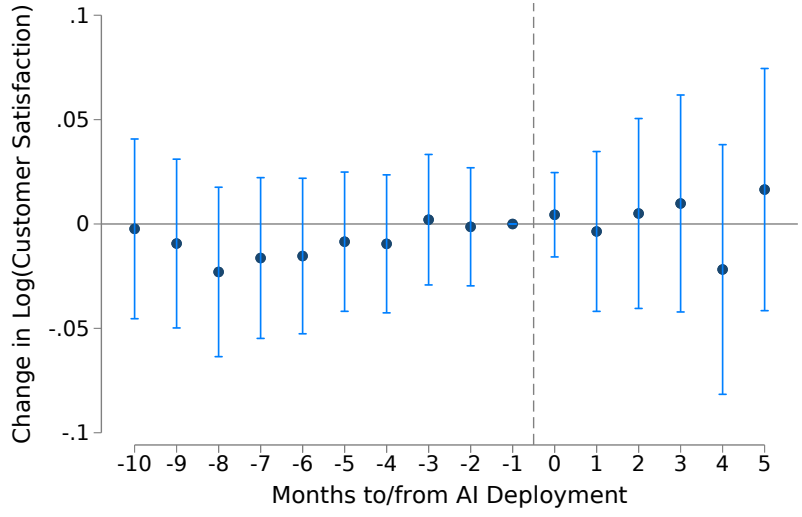


b. Calls per Hour

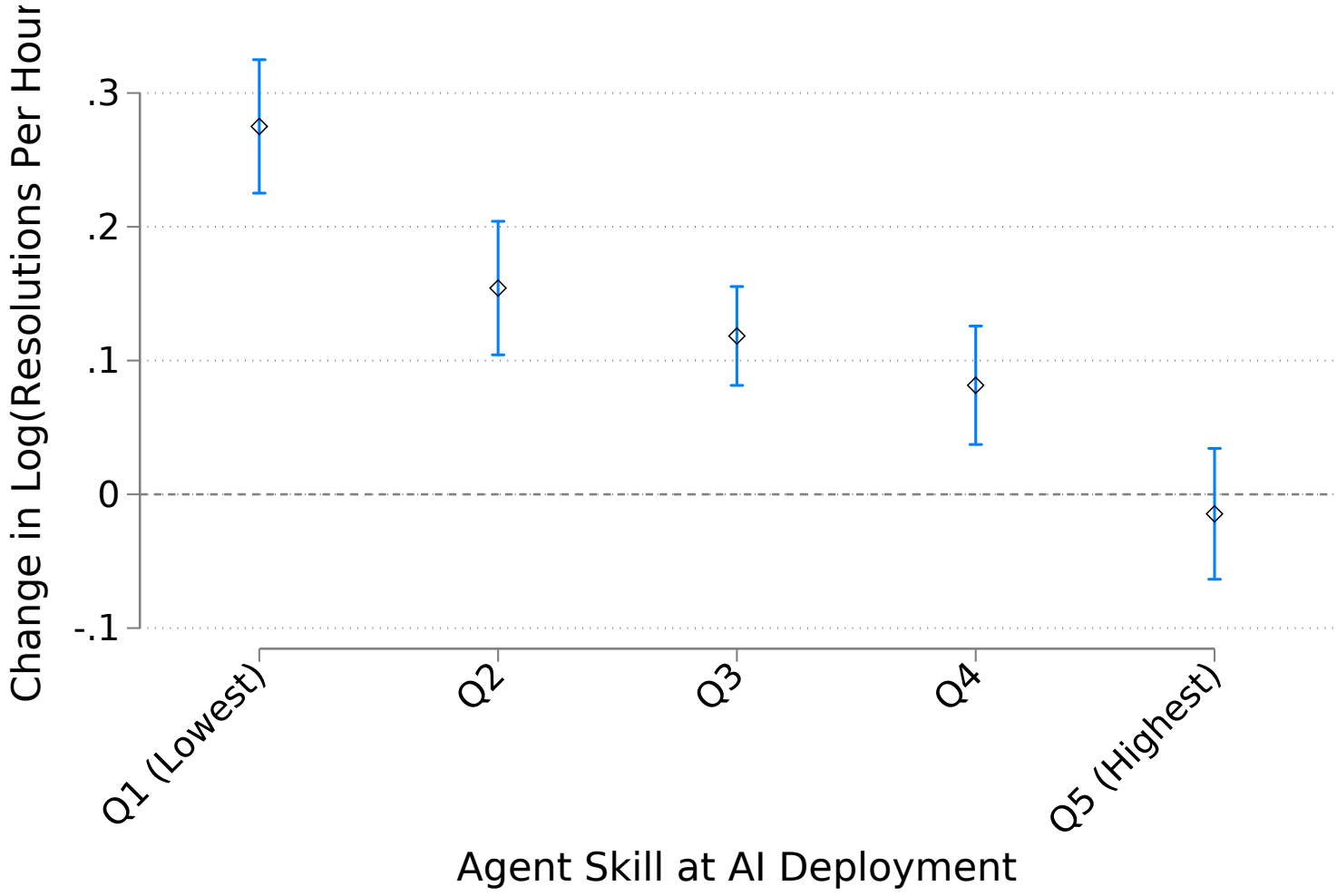


c. Resolution Rate

d. Customer Satisfaction



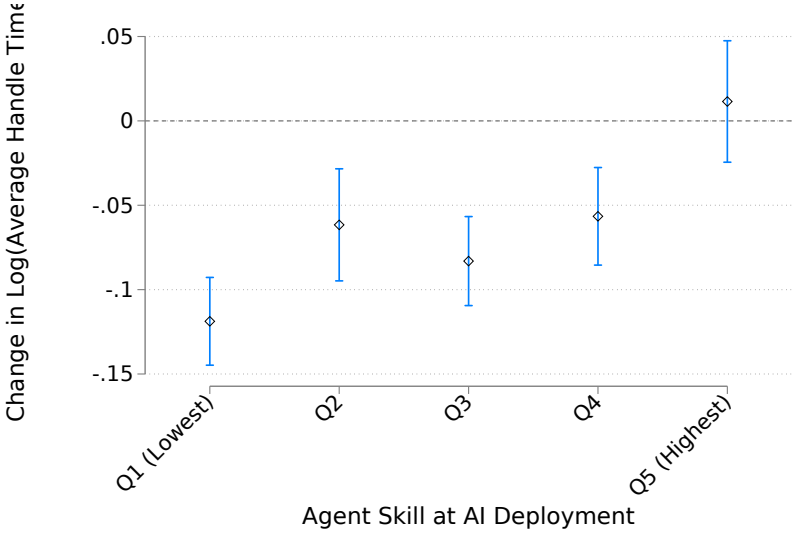
Highest returns for lowest skill agents



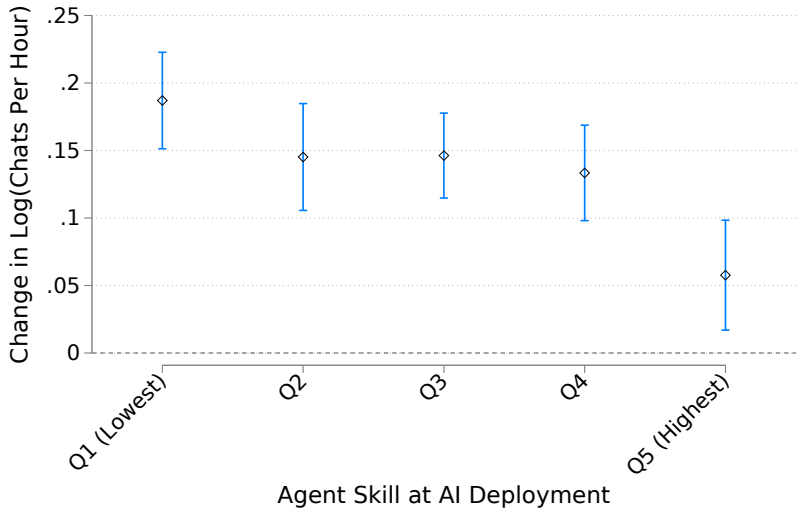
- **Pre-AI agent productivity:** pre-AI index of chats per hour, resolution rate, and customer satisfaction
- Conditional on agent tenure

Other outcomes, by skill

a. Average Handle Time



b. Calls per Hour

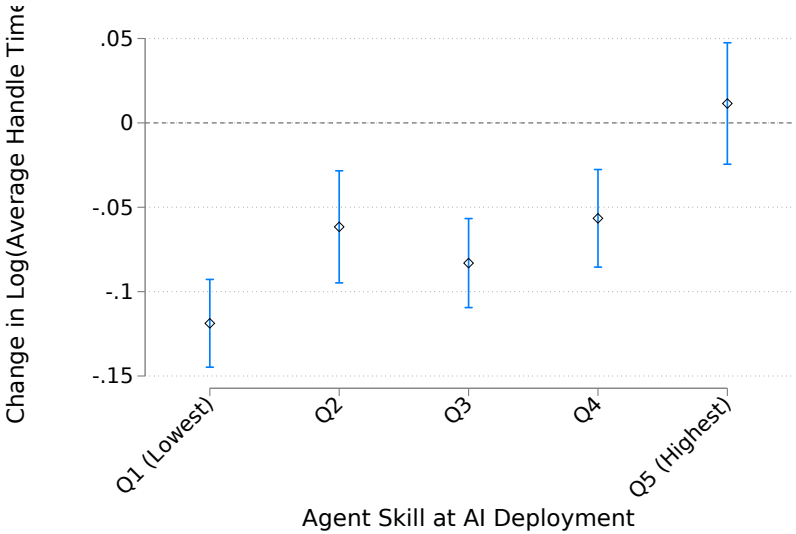


c. Resolution Rate

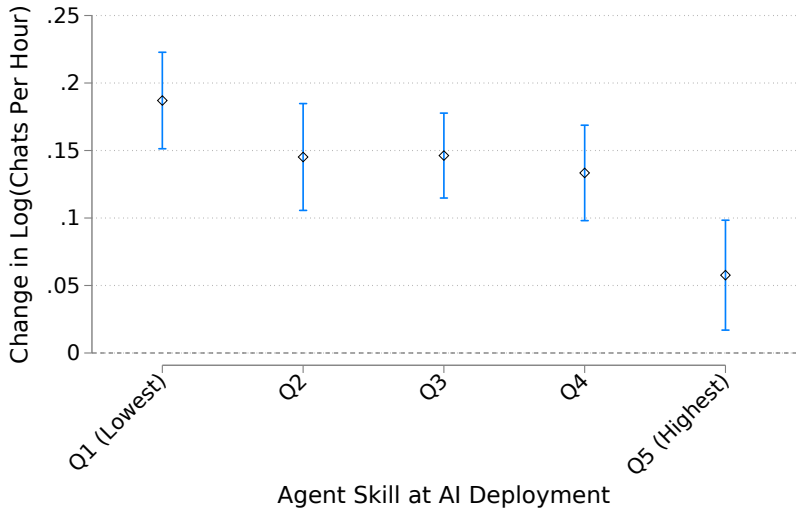
d. Customer Satisfaction

Other outcomes, by skill

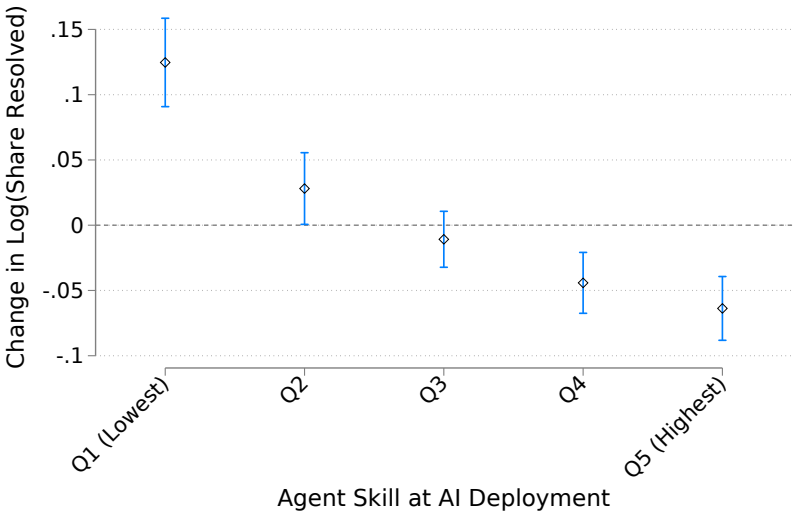
a. Average Handle Time



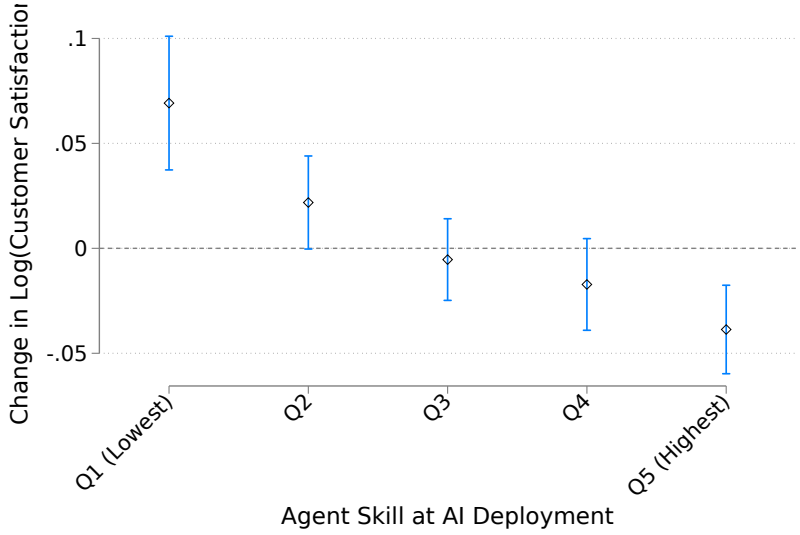
b. Calls per Hour



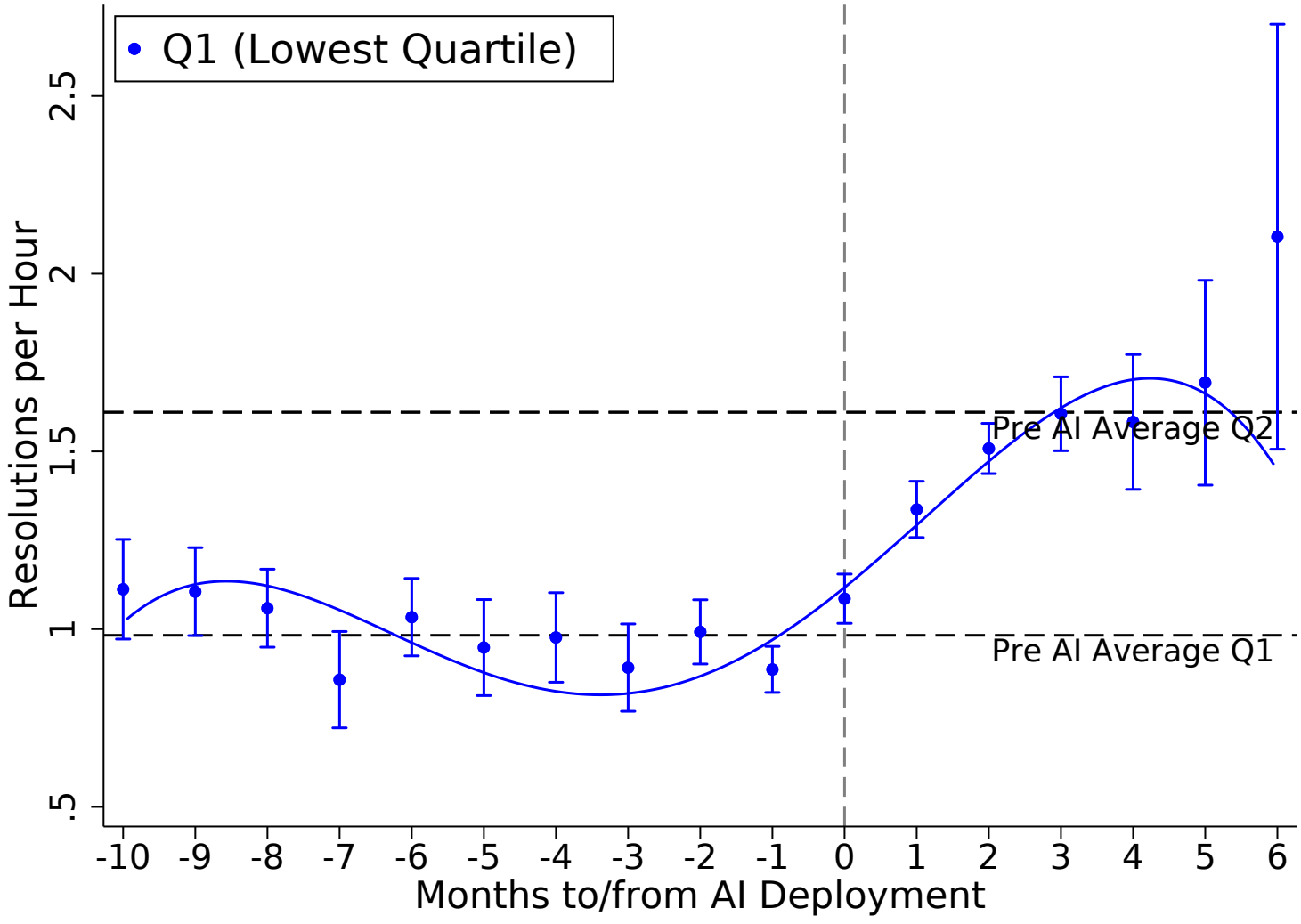
c. Resolution Rate



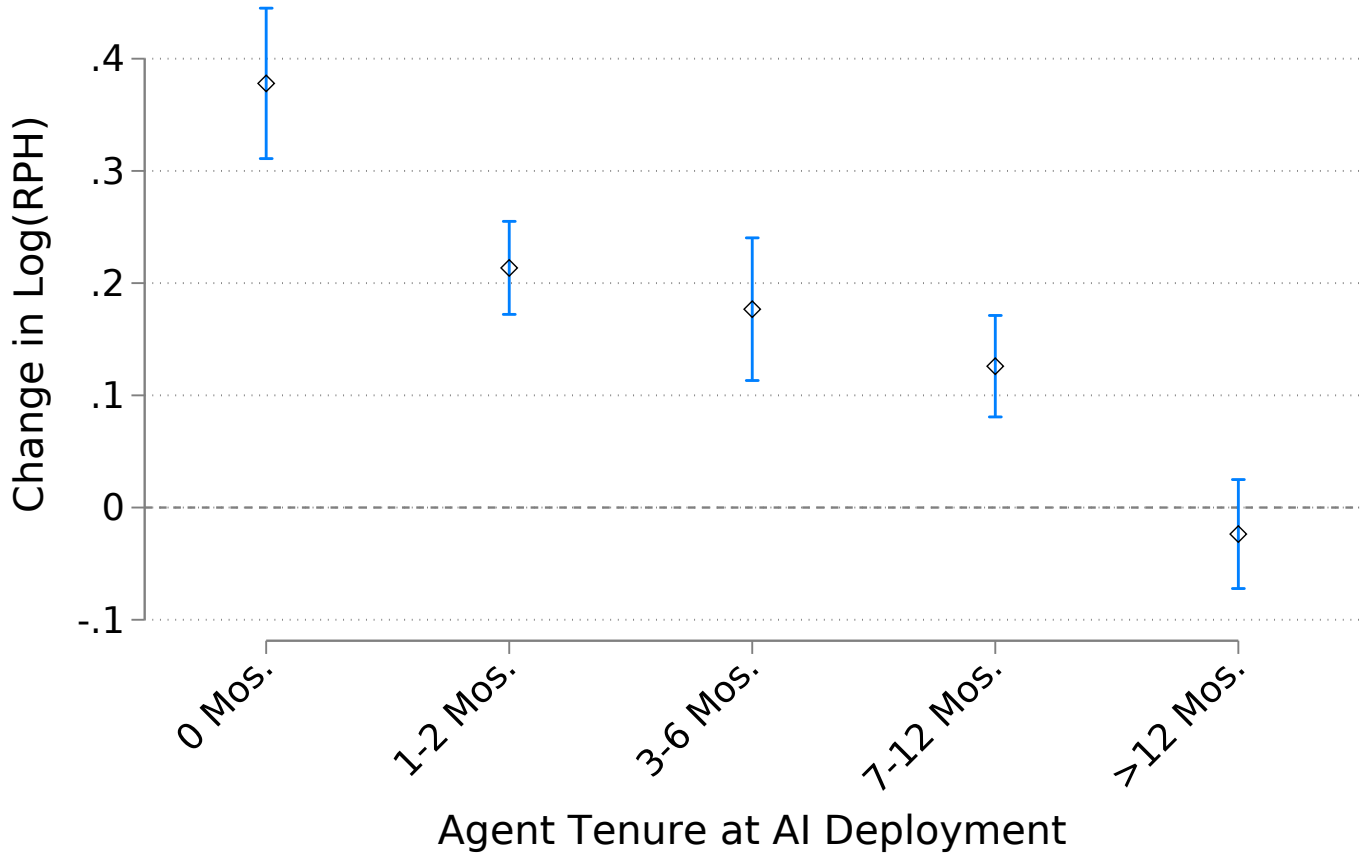
d. Customer Satisfaction



The average agent from the bottom quartile now as productive as the average agent from next quartile



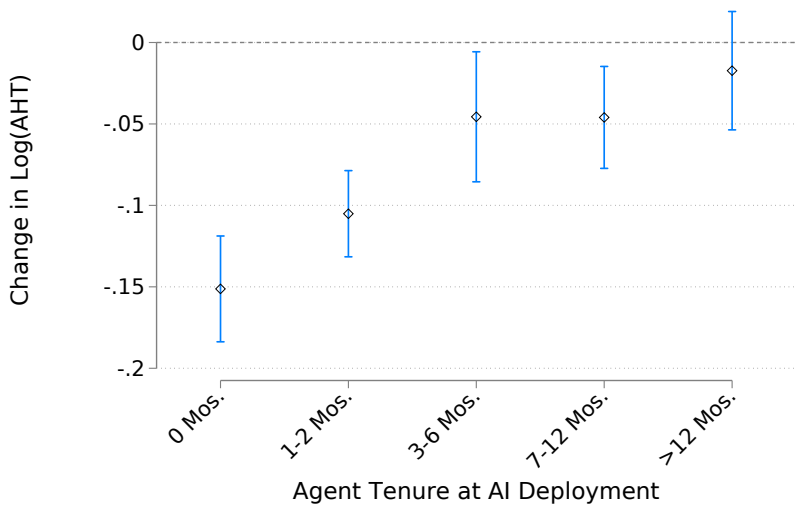
Highest returns for newer agents



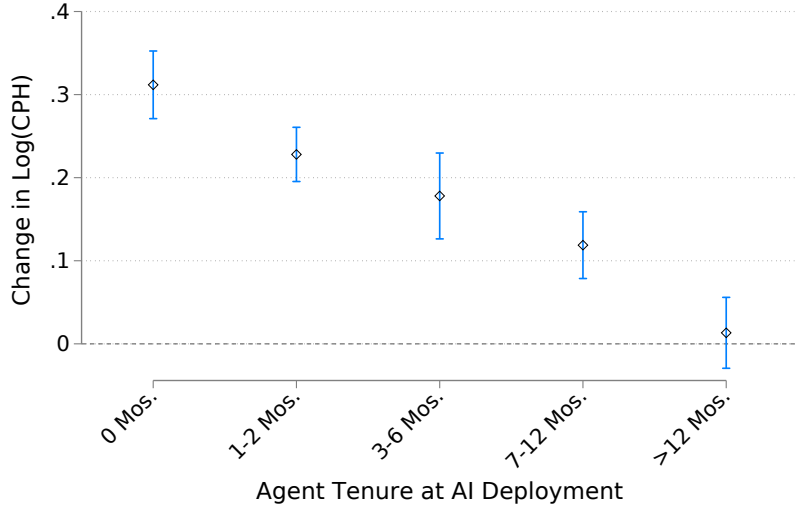
- **Agent tenure:** pre-AI months of experience with the firm
- Results conditional on agent skill, so separate effects with similar pattern

Other outcomes, by tenure

a. Average Handle Time



b. Calls per Hour

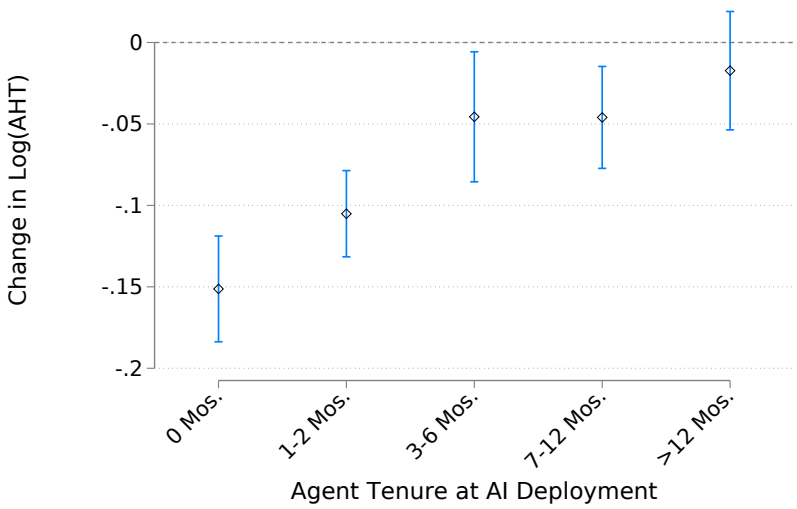


c. Resolution Rate

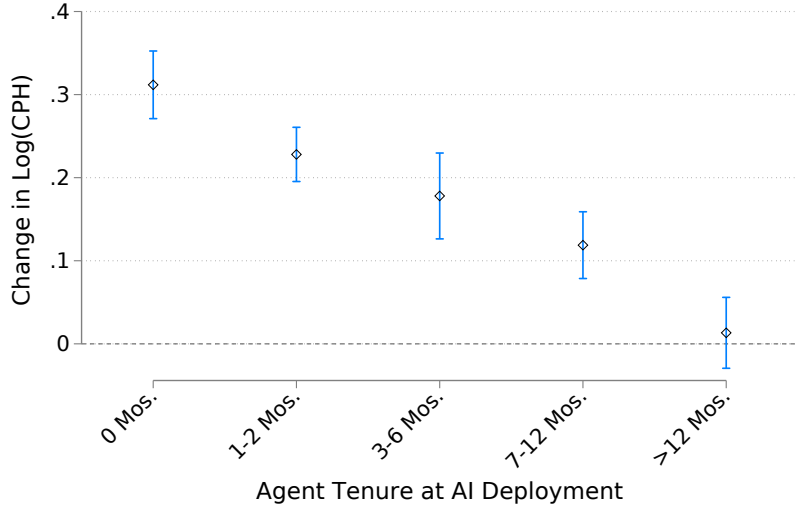
d. Customer Satisfaction

Other outcomes, by tenure

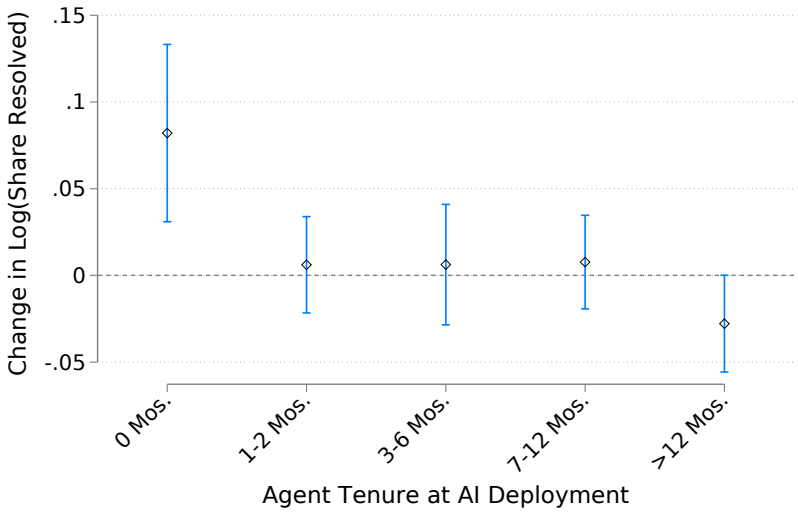
a. Average Handle Time



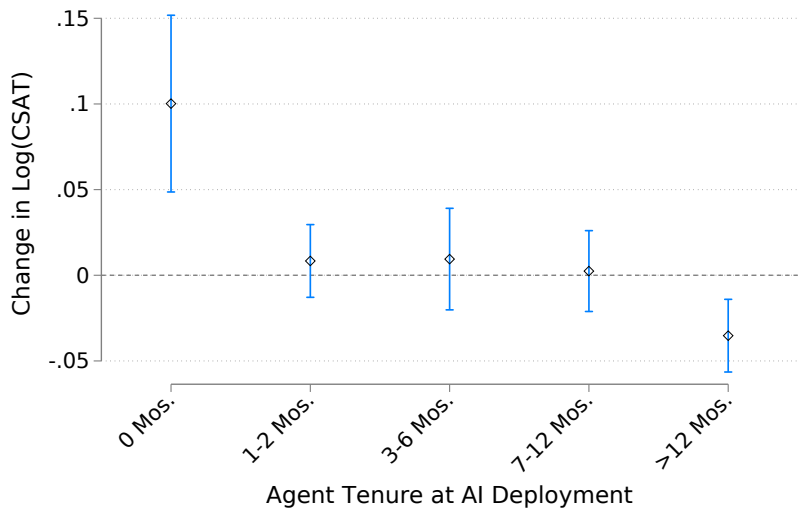
b. Calls per Hour



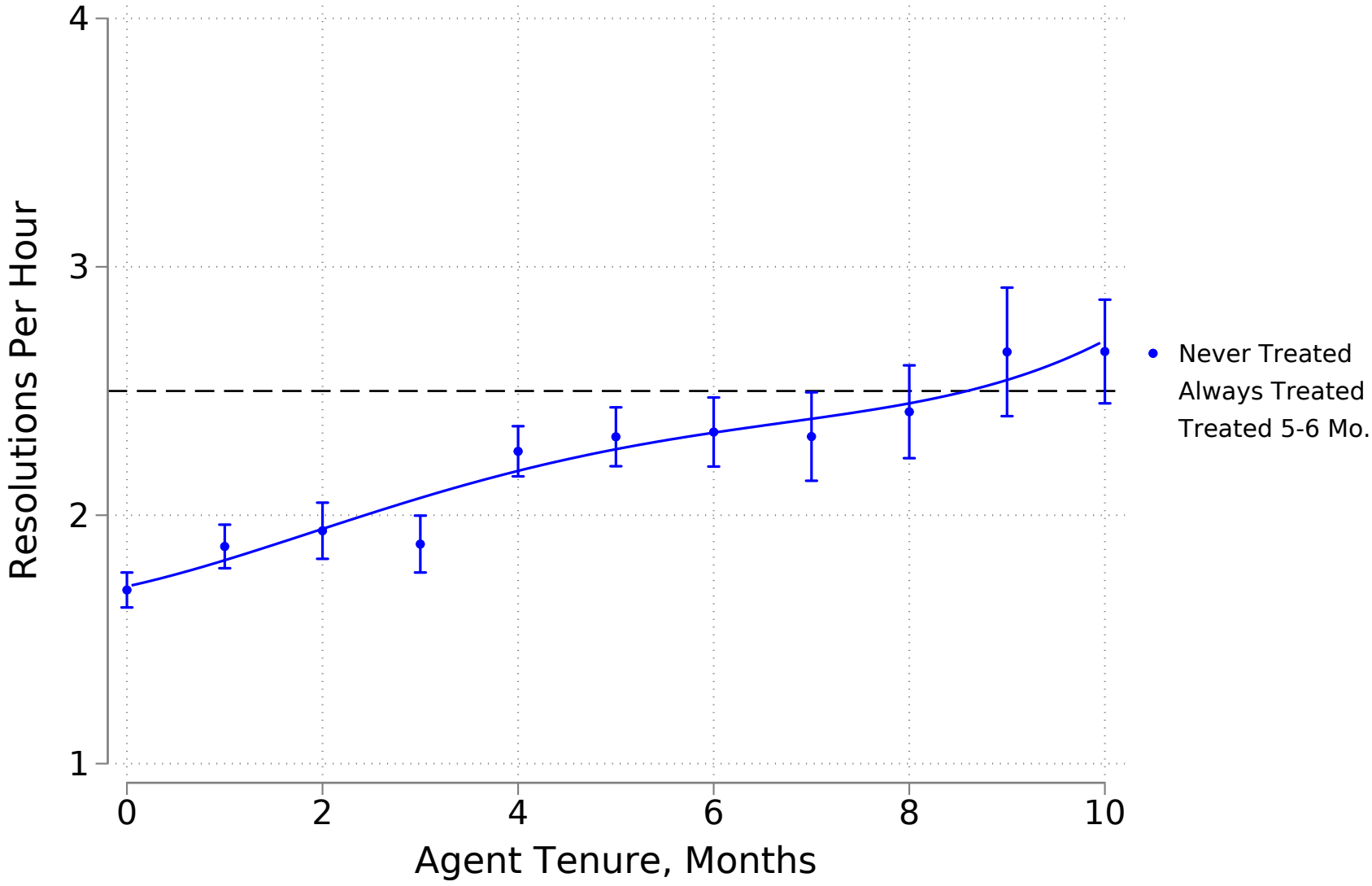
c. Resolution Rate



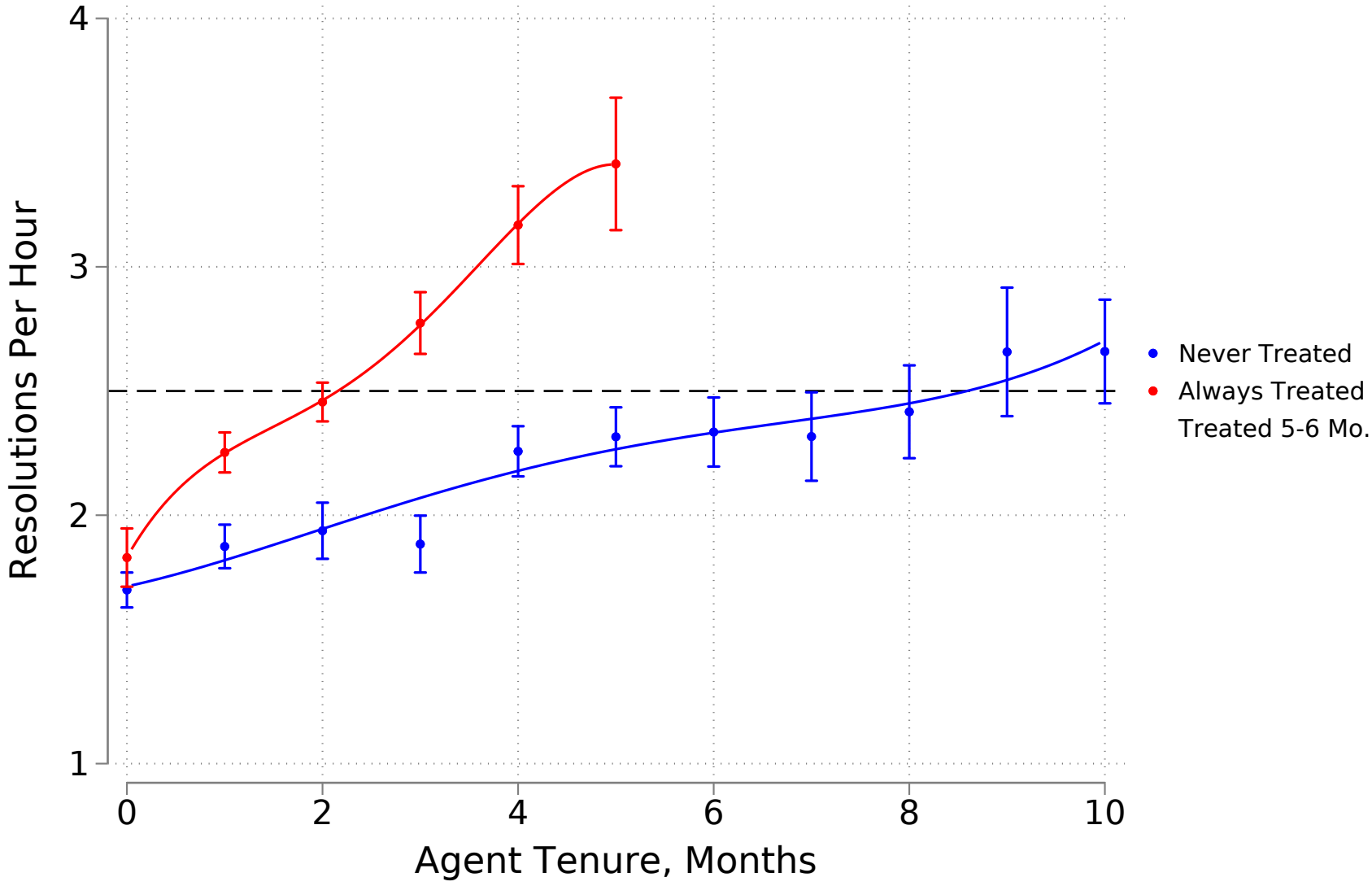
d. Customer Satisfaction



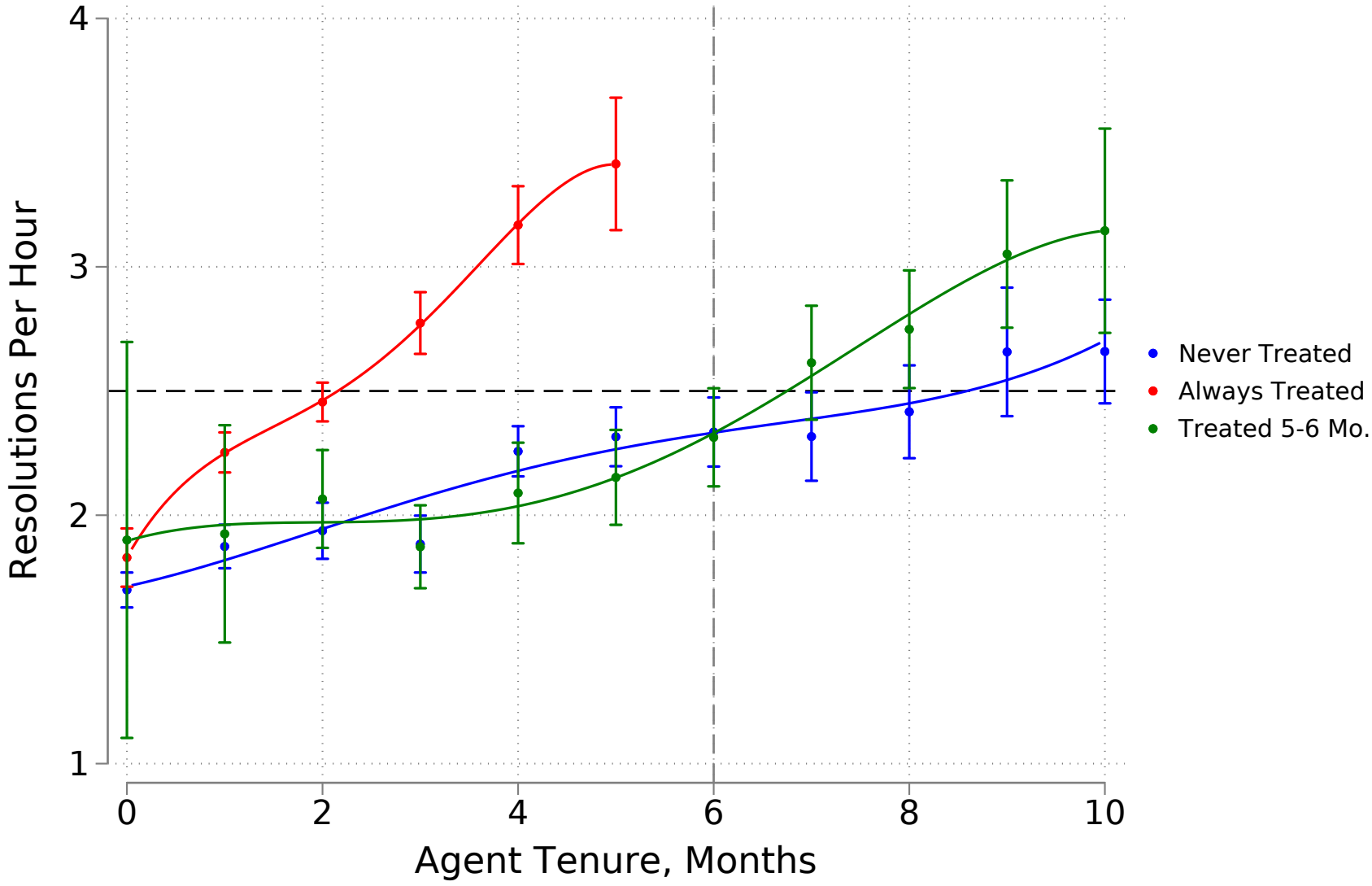
AI assistance helps newer agents “catch up”



AI assistance helps newer agents "catch up"



AI assistance helps newer agents “catch up”



Are workers learning?

- We've shown that worker + AI is more productive than the worker alone

Are workers learning?

- We've shown that worker + AI is more productive than the worker alone

Are workers learning?

- We've shown that worker + AI is more productive than the worker alone

Learning or blind copying?

1. Learning: examples help workers learn how to manage customers and how to solve problems

Are workers learning?

- We've shown that worker + AI is more productive than the worker alone

Learning or blind copying?

1. Learning: examples help workers learn how to manage customers and how to solve problems
2. Blind copying: workers blindly follow AI recommendations and develop less human capital than otherwise

Are workers learning?

- We've shown that worker + AI is more productive than the worker alone

Learning or blind copying?

1. Learning: examples help workers learn how to manage customers and how to solve problems
2. Blind copying: workers blindly follow AI recommendations and develop less human capital than otherwise

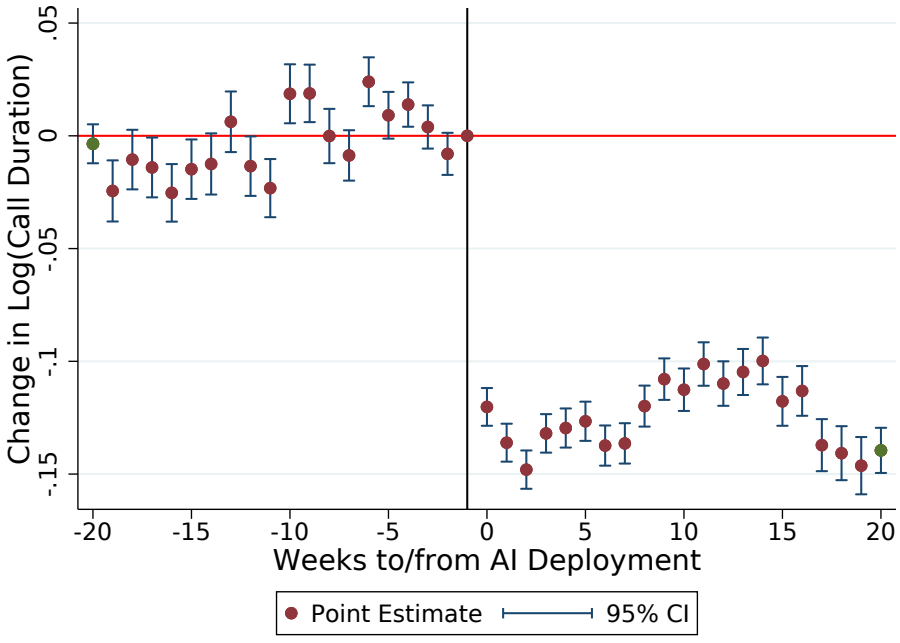
We look at how people are doing in the absence of AI output?

- Outages - periods when AI was not generating output due to software bugs

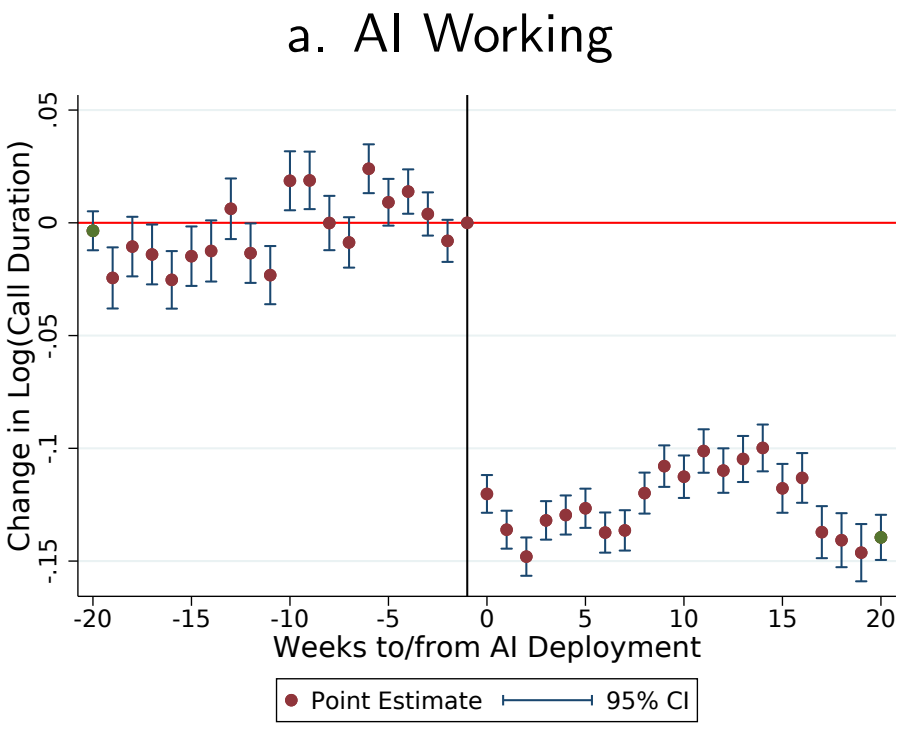
Learning or blind copying?

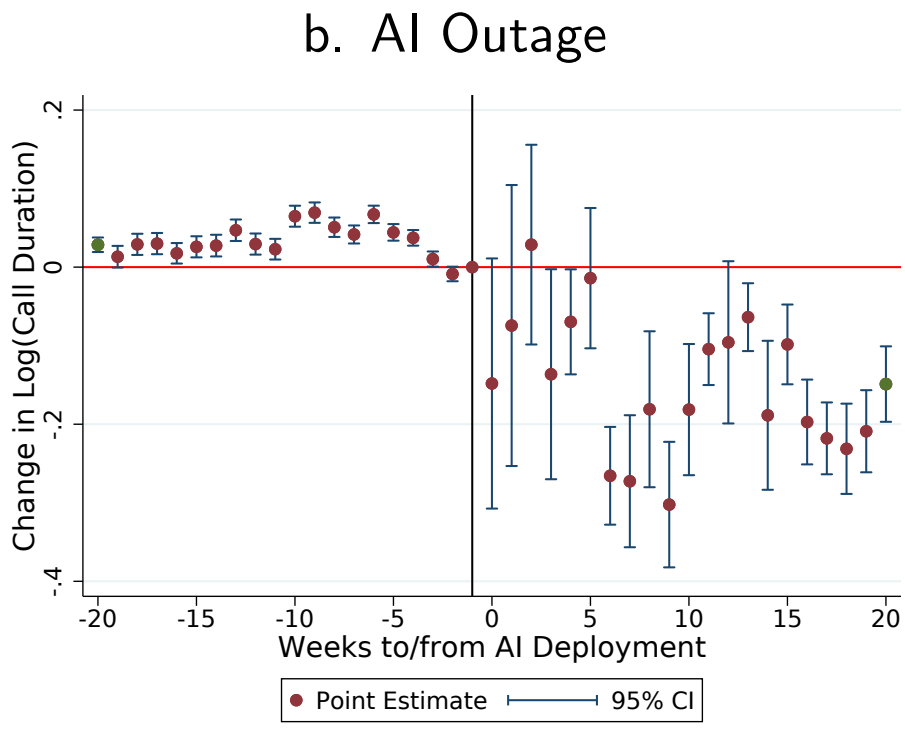
a. AI Working

b. AI Outage



Learning or blind copying?





- AI-enabled workers are still faster during AI outages, but only after about a month of AI exposure

Suggestive textual evidence

We observe complete customer-agent interactions

- Setting is text-based: we observe the entire conversation
- Within agent: full chat history, before and after AI assistance

Suggestive textual evidence

We observe complete customer-agent interactions

- Setting is text-based: we observe the entire conversation
- Within agent: full chat history, before and after AI assistance

Capture language content and tone

- Create semantic text-embedding of chats (using BERT)
- Designed for linguistic comparison

Suggestive textual evidence

We observe complete customer-agent interactions

- Setting is text-based: we observe the entire conversation
- Within agent: full chat history, before and after AI assistance

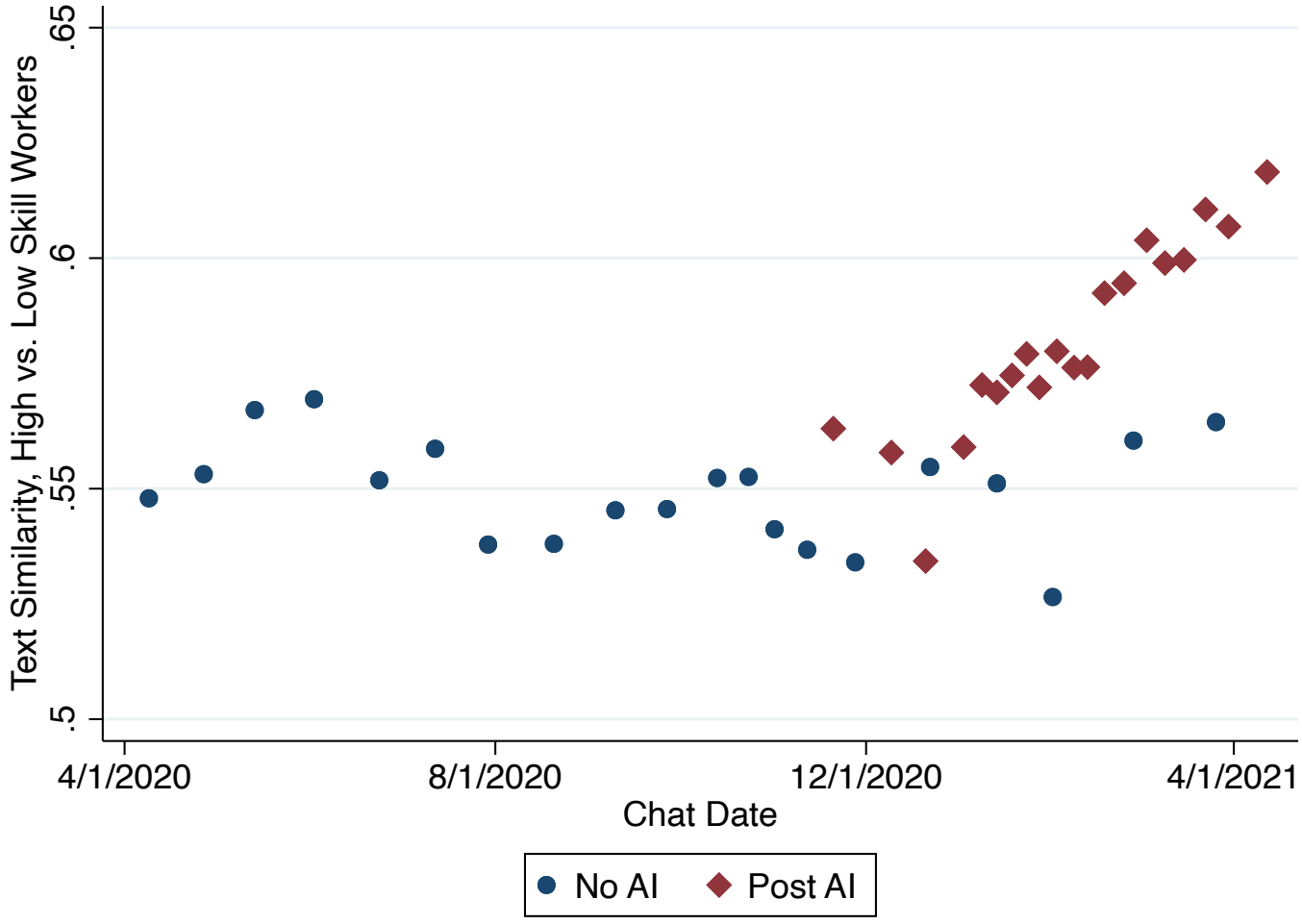
Capture language content and tone

- Create semantic text-embedding of chats (using BERT)
- Designed for linguistic comparison

AI suggestions and "best practices"?

- Does AI assistance help lower-skill workers sound more like higher-skill workers?
  - Within a person, who changes more after AI adoption?
  - Do lower-skill workers sound more like higher-skill workers?

After AI access, low-skill workers sound more like high-skill workers



Low skill workers change more than high skill worker

These are difficult jobs

Customers are often mean...

These are difficult jobs

Customers are often mean...

- Talk with customers about high-stress topics like taxes and payroll

These are difficult jobs

Customers are often mean...

- Talk with customers about high-stress topics like taxes and payroll
- Anonymous encounters—customers not always nice

These are difficult jobs

Customers are often mean...

- Talk with customers about high-stress topics like taxes and payroll
- Anonymous encounters—customers not always nice
- "Are you serious!? You guys are criminals! You are holding money that isn't yours!!! RETURN FUNDS ON CREDIT CARD TRANSACTIONS IMMEDIATELY!"

These are difficult jobs

Customers are often mean...

- Talk with customers about high-stress topics like taxes and payroll
- Anonymous encounters—customers not always nice
- "Are you serious!? You guys are criminals! You are holding money that isn't yours!!! RETURN FUNDS ON CREDIT CARD TRANSACTIONS IMMEDIATELY!"
- Managing this is a **skill**:
  - "But I always tell people we can only catch what they throw, so if they throw a baseball, we can't catch a football.."

These are difficult jobs

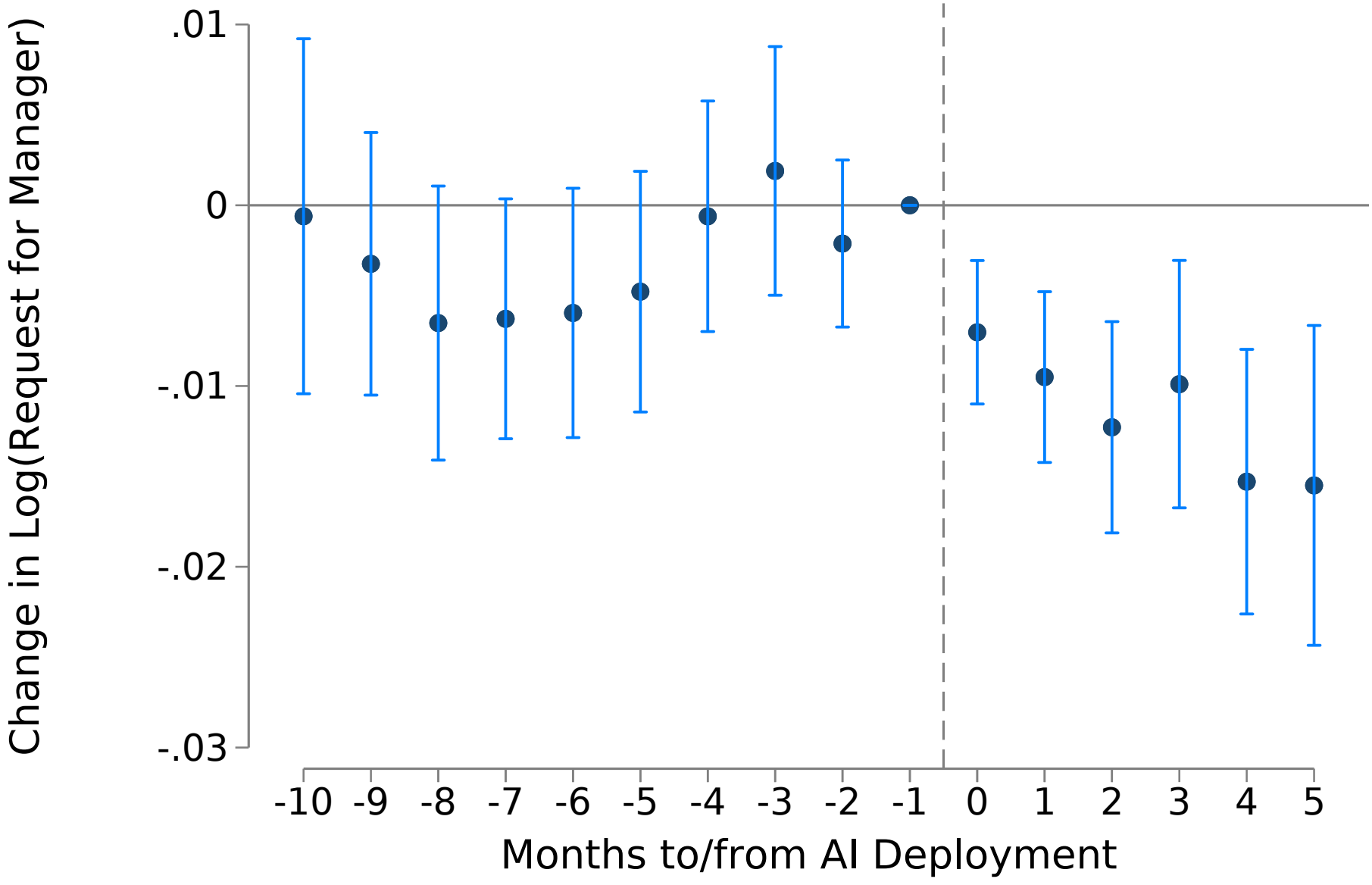
Customers are often mean...

- Talk with customers about high-stress topics like taxes and payroll
- Anonymous encounters—customers not always nice
- "Are you serious!? You guys are criminals! You are holding money that isn't yours!!! RETURN FUNDS ON CREDIT CARD TRANSACTIONS IMMEDIATELY!"
- Managing this is a **skill**:
  - "But I always tell people we can only catch what they throw, so if they throw a baseball, we can't catch a football.."

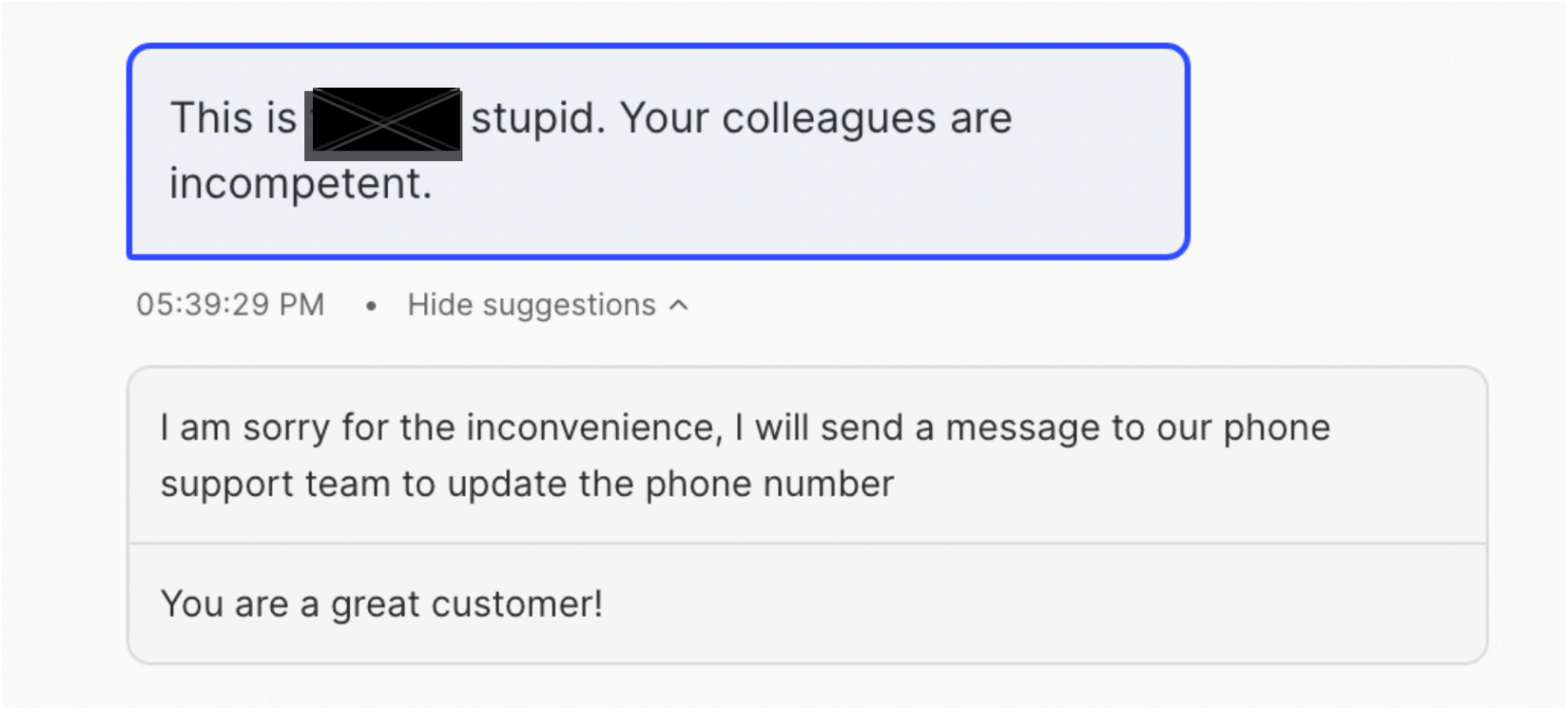
This impacts the experience of work

- Industry-wide, estimated that 60% of workforce turns over in a year

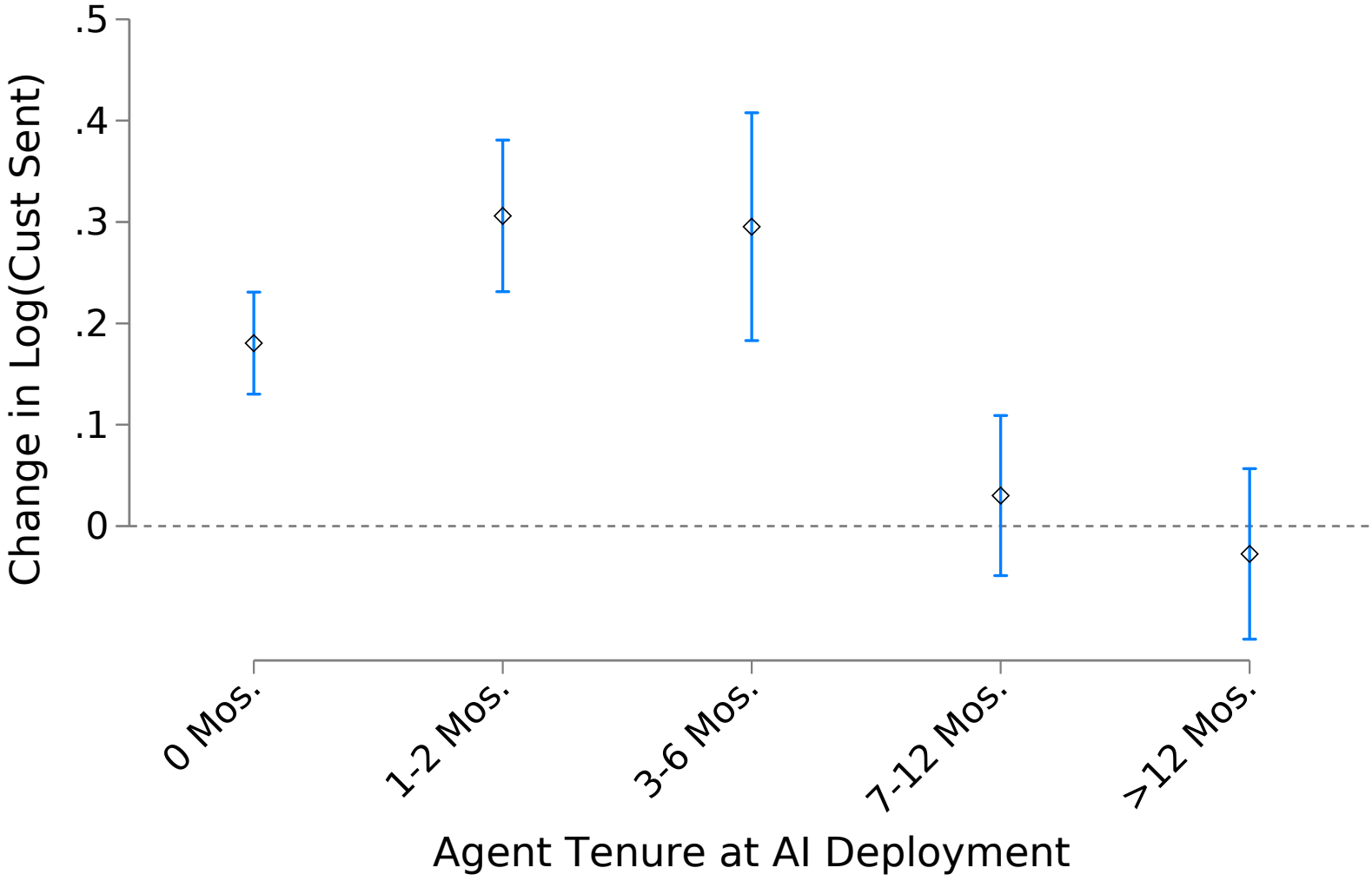
Customers demand to speak with a manager less frequently



AI assistance

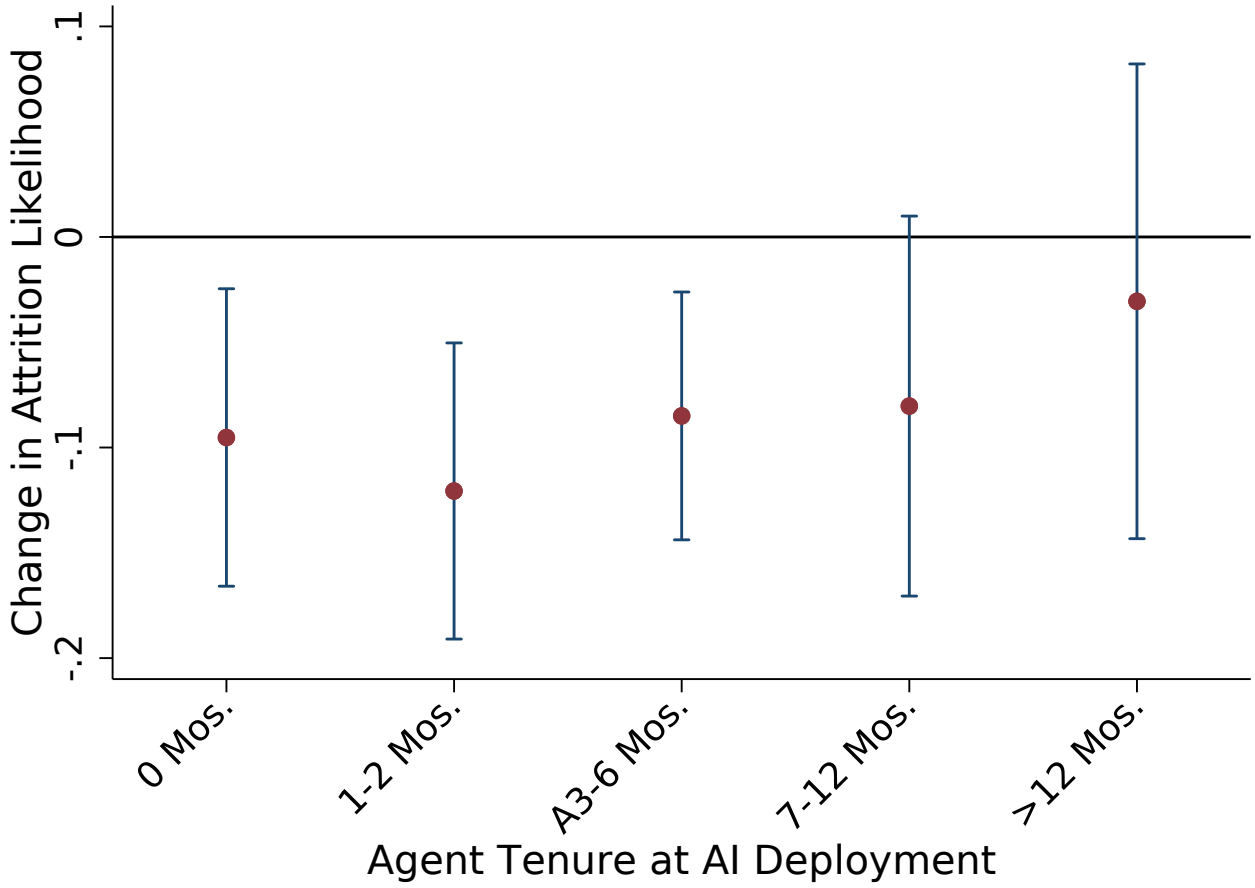


Less mean customers - particularly for least experienced workers



No change in agent tone

Evidence that AI access reduces turnover



- Caveat: rollout happens at the agent level and agents can only leave once, so could overstate these effects if AI is less likely to be assigned to agents about to leave.

Conclusion

Generative AI increases productivity, mostly for less experienced and less able workers

- Improvements in efficiency and quality measures
- Some evidence of improvements in the experience of work
- Driven by the newest and least skilled workers

Implications

- Look for impacts of AI in new places
  - Changing value of experience, managers, worker training?
- Raises new questions;
  - Incentivizing and compensating workers for these "best practices"?

Questions: lraymond@mit.edu

Appendix

Summary statistics

Variable	All	Never Treated	Treated, Pre	Treated, Post
Chats	3,007,501	945,954	882,105	1,180,446
Agents	5,179	3,523	1,341	1,636
Number of Teams	133	111	80	81
Share US Agents	.11	.15	.081	.072
Distinct Locations	25	25	18	17
Average Chats per Month	127	83	147	188
Average Handle Time (Min)	41	43	43	35
St. Average Handle Time (Min)	23	24	24	22
Resolution Rate	.82	.78	.82	.84
Resolutions Per Hour	2.1	1.7	2	2.5
Customer Satisfaction (NPS)	79	78	80	80

Table: Sample Summary Statistics

Study Design

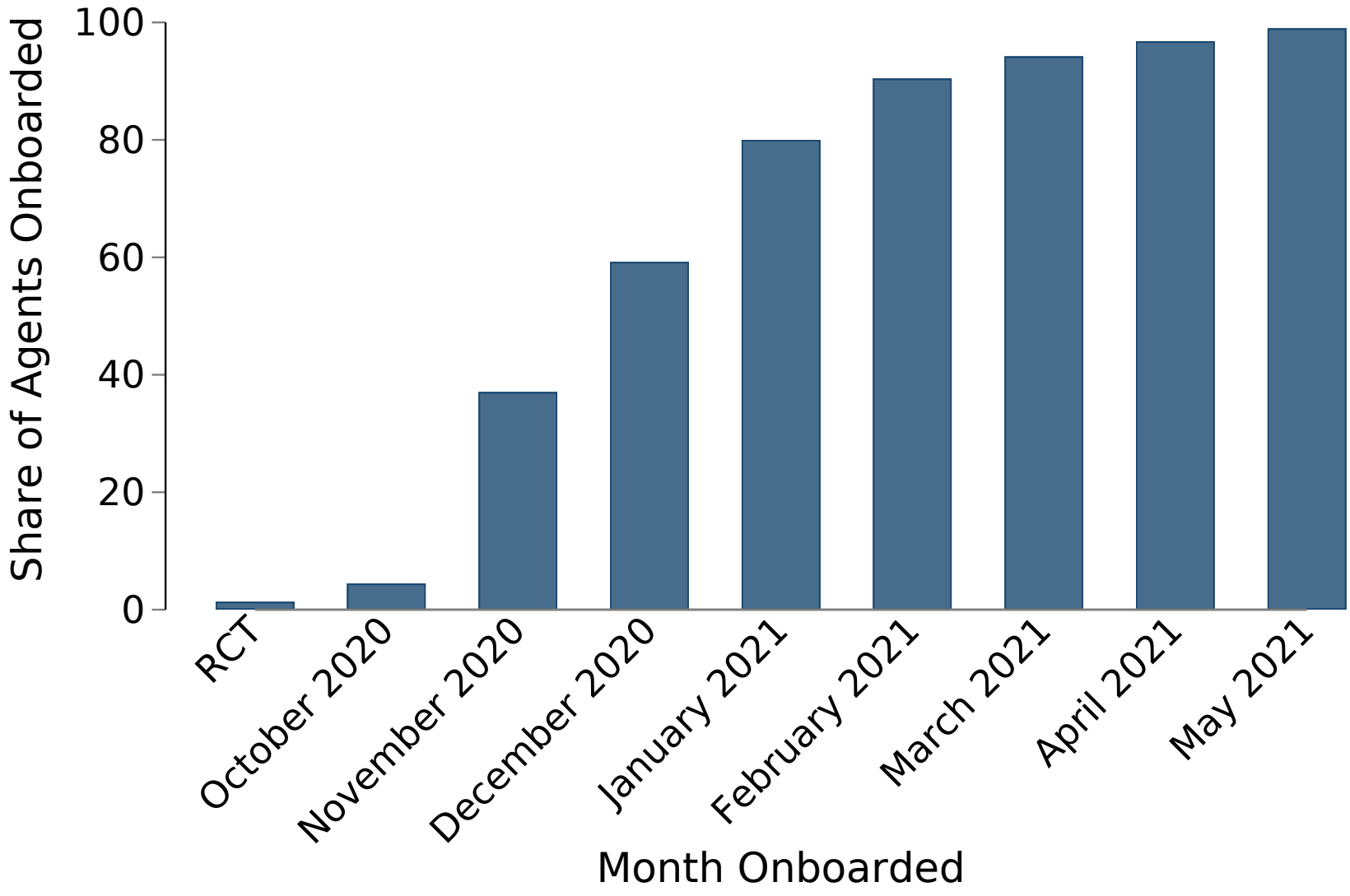
Balance table

Variable	(1) Control	(2) Treatment Pre-AI	(3) Difference
AHT (Minutes)	35.782	35.311	-0.471***
	(13.503)	(13.470)	(0.000)
Resolution Rate	0.800	0.833	0.033***
	(0.203)	(0.147)	(0.000)
Customer Satisfaction	75.561	78.673	3.112***
	(19.885)	(14.351)	(0.000)
Res. per Hour	2.095	2.256	0.160***
	(0.809)	(0.703)	(0.000)
Observations	946,056	882,105	1,828,161

Table: Pre-Treatment Balance Table

Study Design

Deployment timeline



AI model deployment raises productivity

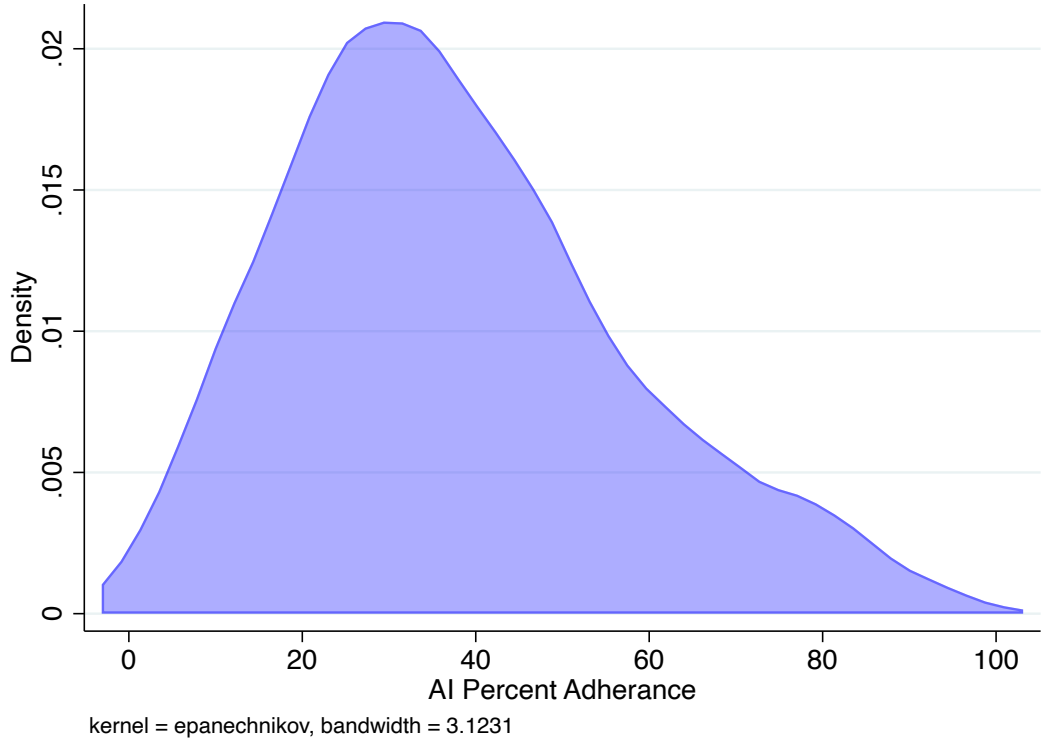
VARIABLES	(1) Res./Hr	(2) Res./Hr	(3) Res./Hr	(4) Log(Res./Hr)	(5) Log(Res./Hr)	(6) Log(Res./Hr)
Post AI X Ever Treated	0.469***	0.371***	0.301***	0.221***	0.180***	0.138***
	(0.0542)	(0.0519)	(0.0498)	(0.0211)	(0.0188)	(0.0199)
Ever Treated	0.110*			0.0581*		
	(0.0589)			(0.0321)		
Observations	13,192	12,295	12,295	12,747	11,875	11,875
R-squared	0.249	0.562	0.575	0.260	0.572	0.593
Year Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Location FE	Yes	Yes	Yes	Yes	Yes	Yes
Agent FE	-	Yes	Yes	-	Yes	Yes
Agent Tenure	-	-	Yes	-	-	Yes
DV Mean	2.123	2.176	2.176	0.709	0.734	0.734

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.10

- Preferred specification in Columns 3 and 6
- On average, access to AI increases agent productivity by .30 resolutions per hour or 14 percent

Event Study

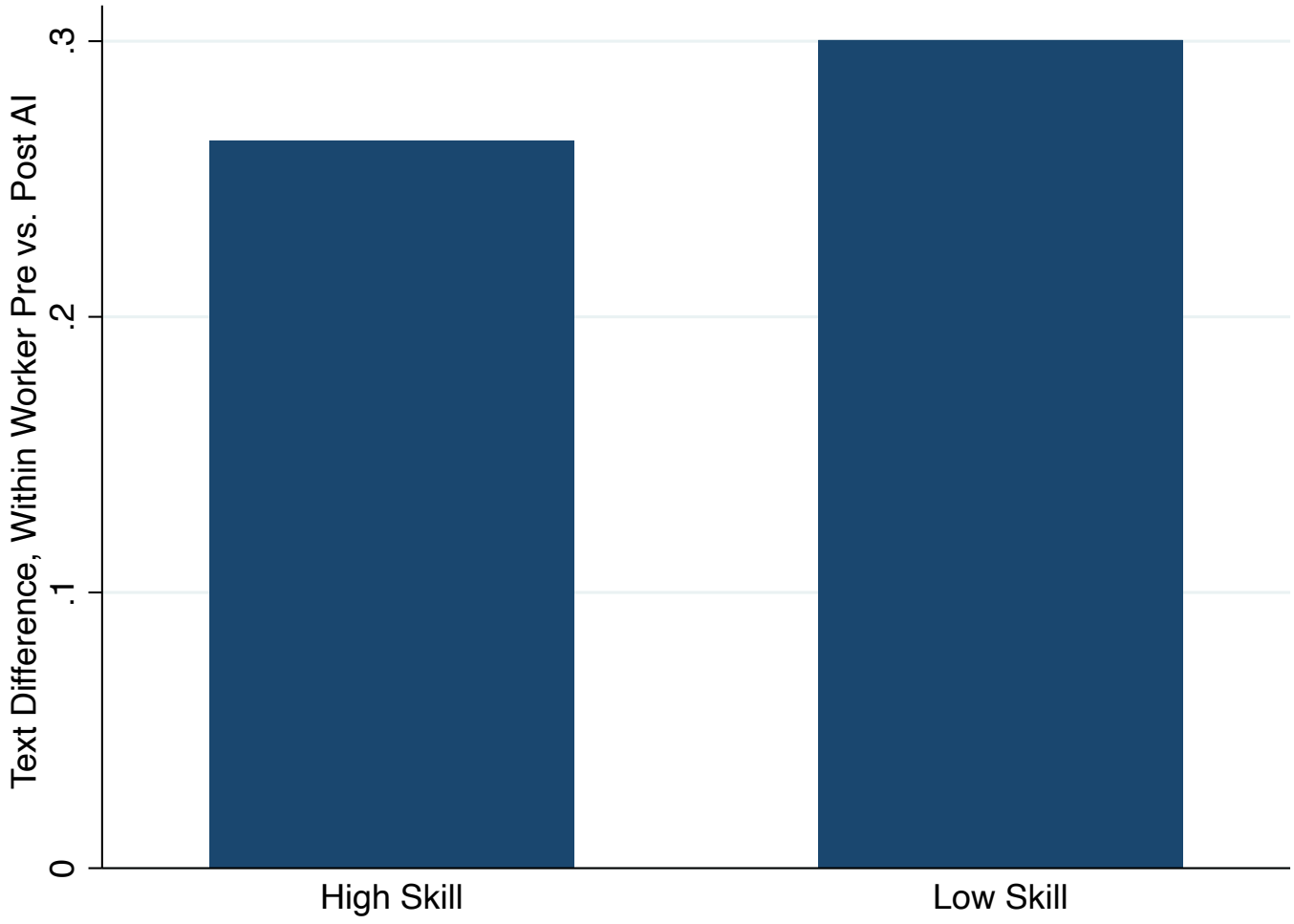
Agents follow about 35 percent of recommendations



- Sometimes. Agents are not blindly copying recommendations
- **Adherence:** the share of AI recommendations agent adopts (either by pasting directly or writing something very similar)

Event Study

Lower skill workers' language changes more



Skill gap text similarity

No change in agents tone