Augmented Intelligence: The Effects of AI on Productivity and Work Practices

Lindsey Raymond

Joint work with Erik Brynjolfsson and Danielle Li
September 2022

What exactly is ML good at?

Improving worker performance is hard

- Many examples of large performance differences
- Good workers are often distinguished by tacit skills that are difficult to transfer: how do you tell an RA to write a good introduction?

What exactly is ML good at?

Improving worker performance is hard

- Many examples of large performance differences
- Good workers are often distinguished by tacit skills that are difficult to transfer: how do you tell an RA to write a good introduction?

Can ML help?

What exactly is ML good at?

Improving worker performance is hard

- Many examples of large performance differences
- Good workers are often distinguished by tacit skills that are difficult to transfer: how do you tell an RA to write a good introduction?

Can ML help?

ML is a powerful tool for making distinctions

- Discovers what separates good outcomes from bad outcomes: medical diagnoses, hiring and lending decisions etc.

What exactly is ML good at?

Improving worker performance is hard

- Many examples of large performance differences
- Good workers are often distinguished by tacit skills that are difficult to transfer: how do you tell an RA to write a good introduction?

Can ML help?

ML is a powerful tool for making distinctions

- Discovers what separates good outcomes from bad outcomes: medical diagnoses, hiring and lending decisions etc.
- Training data is generated by humans: the actions of doctors, managers, loan officers etc.

What exactly is ML good at?

Improving worker performance is hard

- Many examples of large performance differences
- Good workers are often distinguished by tacit skills that are difficult to transfer: how do you tell an RA to write a good introduction?

Can ML help?

ML is a powerful tool for making distinctions

- Discovers what separates good outcomes from bad outcomes: medical diagnoses, hiring and lending decisions etc.
- Training data is generated by humans: the actions of doctors, managers, loan officers etc.
- ML makes distinctions on human generated data, and can distinguish good from bad

This paper: ML and workplace productivity

How might ML be useful?

- Identify good people

This paper: ML and workplace productivity

How might ML be useful?

- Identify good people
- But ML can do more than just identify; it can predict how effective workers will **behave**

This paper: ML and workplace productivity

How might ML be useful?

- Identify good people
- But ML can do more than just identify; it can predict how effective workers will **behave**
- Learns what makes good people good–not just who they are

This paper: ML and workplace productivity

How might ML be useful?

- Identify good people
- But ML can do more than just identify; it can predict how effective workers will **behave**
- Learns what makes good people good–not just who they are

ML and knowledge transfer

- Learn successful behaviors (even tacit) from top performers
- Turn behaviors into complex decision rules
- Share these suggestions with all workers

This paper: ML and workplace productivity

How might ML be useful?

- Identify good people
- But ML can do more than just identify; it can predict how effective workers will **behave**
- Learns what makes good people good–not just who they are

ML and knowledge transfer

- Learn successful behaviors (even tacit) from top performers
- Turn behaviors into complex decision rules
- Share these suggestions with all workers

Our paper: can ML augment human decision making?

Can ML augment human decision making?

Questions

1. Can suggestion-based augmentation impact worker productivity?
2. What sorts of things can the ML learn and which types of workers are impacted most, and how?
3. What are the impacts on the organization and experience of work?

Can ML augment human decision making?

Questions

1. Can suggestion-based augmentation impact worker productivity?
2. What sorts of things can the ML learn and which types of workers are impacted most, and how?
3. What are the impacts on the organization and experience of work?

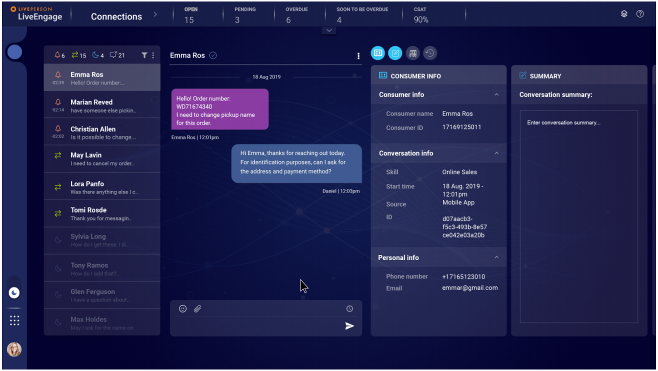
Setting

- Staggered rollout of ML in Fortune 500 chat based technical support
- Random allocation of problems, large productivity differences
- Empirical features that allow us to get at knowledge transfer
- Human + ML interactions "in the wild"

Our findings

1. Overall effects:
   - **Increases in average productivity**
   - **Humans continues to perform best on most uncommon problems**
2. Distributional consequences:
   - **Increases driven by ex-ante less skilled, less experienced and outsourced and international workers**
3. Organizational implications:
   - **Changes in communication patterns, manager help, team size**

Customer support is one of the top use cases for ML



- Technical feasibility: Advancements in NLP and reinforcement learning applied to huge amount of customer-agent conversations, labeled with outcomes (successful resolution, customer satisfaction)
- Business need: High turnover jobs (2-3 months in training)

Why is customer support an ML problem?

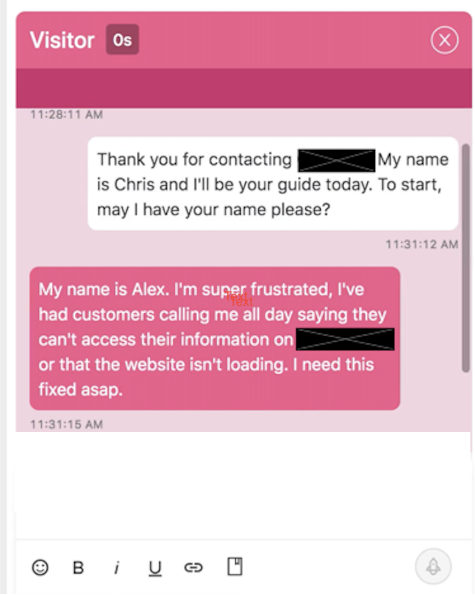
A series of pattern recognition exercises

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?

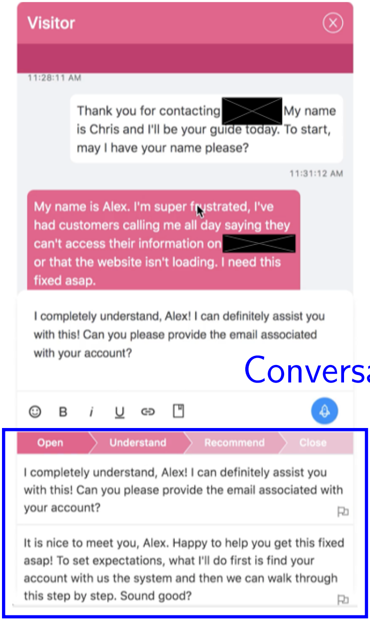
Key insight: ML can encode even tacit skills

- ML recognizes differences between top agents and everyone else
- Successful behaviors are turned into real-time suggestions without anyone needing to understand why they work.

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
B
i
U

Suggested Responses



Conversational Responses

Data and study design

Data

- Conversations: Four millions chats between 2019-2022
- Agents: 3,000 agents, 140 teams and 5 firms
- Data: Chat text, topic, content, AI output, agent interactions
- Outcomes: Technical issues solved per hour and customer satisfaction

Study Design

- AI rolled out across agents over time Timeline

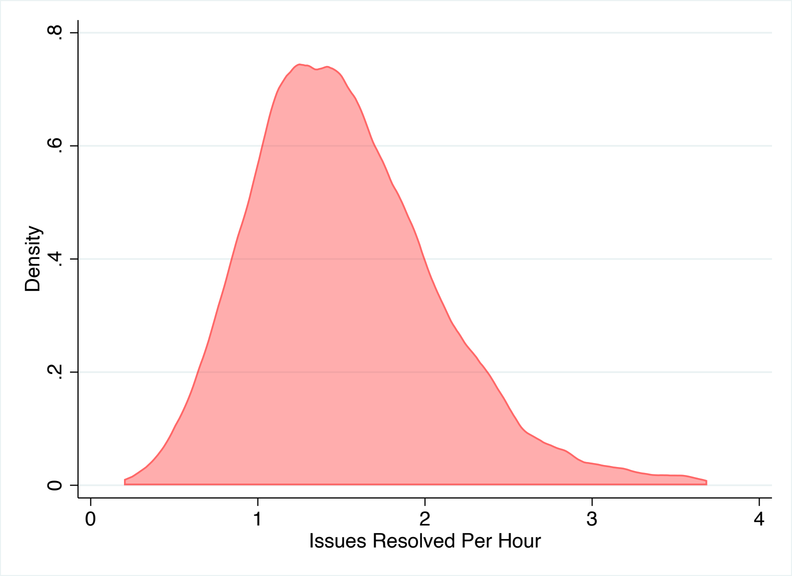
$$Y_{it} = \alpha_i + \gamma_t + \beta_1 (Treated_i x Post_{it}) + \epsilon_{it}$$

- Analyze using a stacked event study design Specifying equation
- Observations aggregated to the agent-week or agent-month level
- Text analysis done with BERT Text analysis details

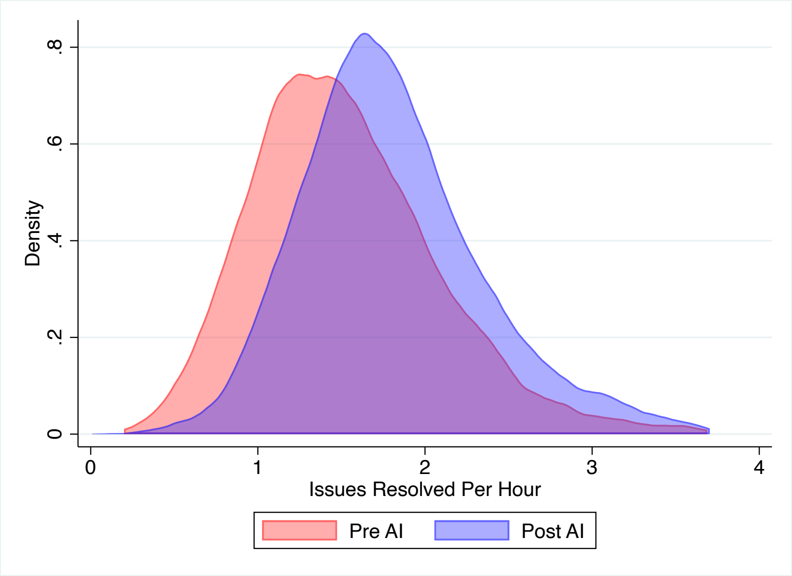
Results Overview

1. Overall agent-level productivity impact
2. By agent skill
3. By agent tenure
4. By agent-firm relationship
5. By problem type

Productivity gains are evident in the raw data



Productivity gains are evident in the raw data

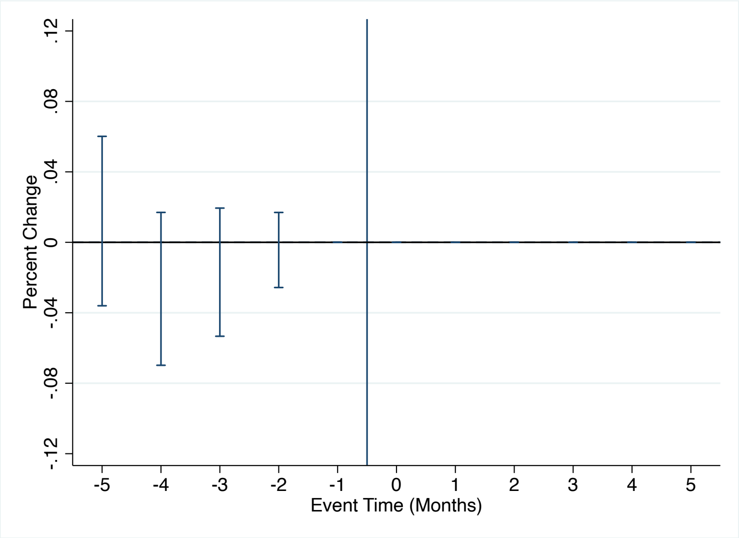


ML deployment raises productivity

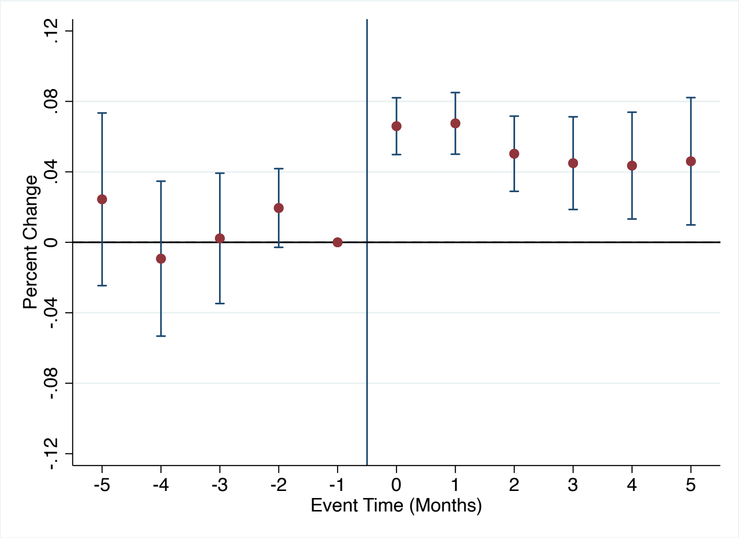
	(1) Productivity	(2) Share Resolved	(3) Calls per Hour
ML Deployment	0.070***	0.063*	0.072***
	(0.008)	(0.032)	(0.008)
Unit FE	Agent x Exp	Agent x Exp	Agent x Exp
Time FE	Month x Exp	Month x Exp	Month x Exp
Observations	60,856	50,704	60,856
Pre Mean	1.5	81.75	1.61

- Agent gains a month of experience
- Eight month median worker tenure with 2-3 months in training
- Also consistency effects

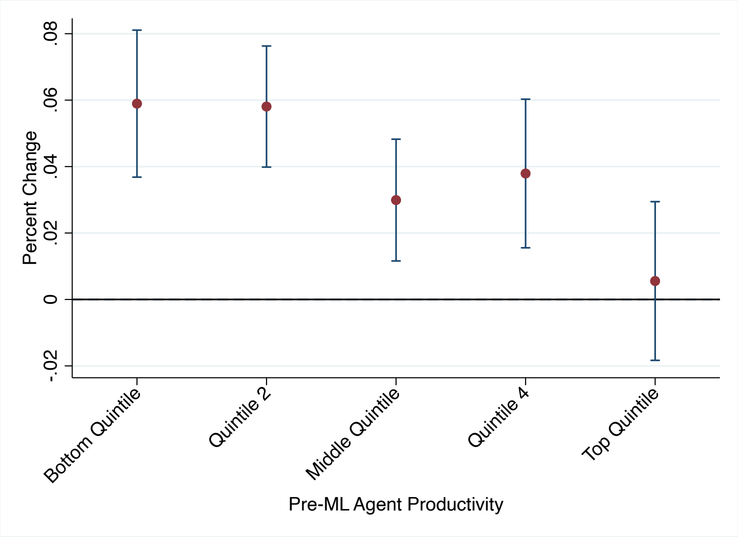
Productivity improvements persist



Productivity improvements persist

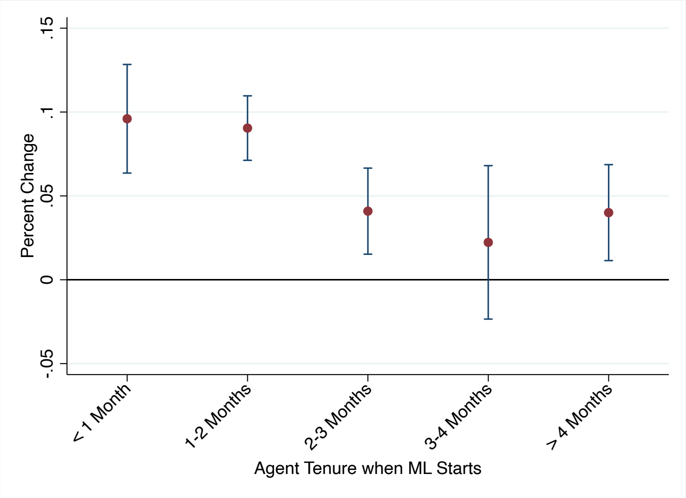


Highest returns for lowest skill agents



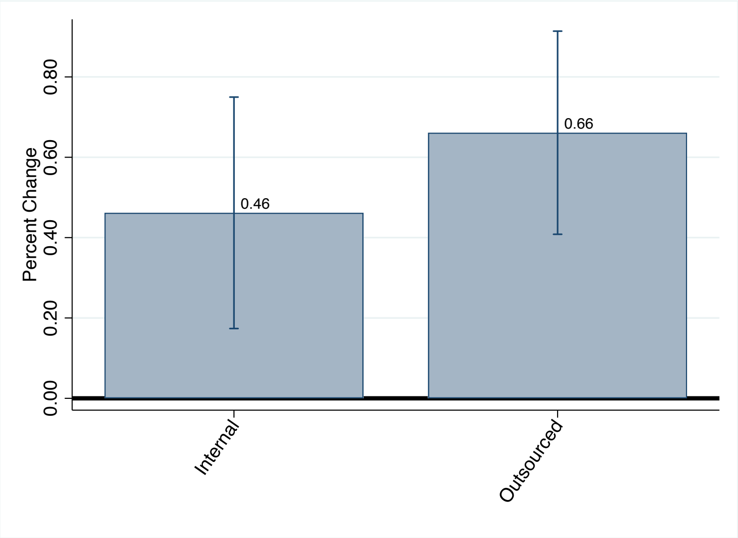
- **Pre-ML agent productivity:** agent's relative performance within firm prior to the ML

Highest returns for newer agents

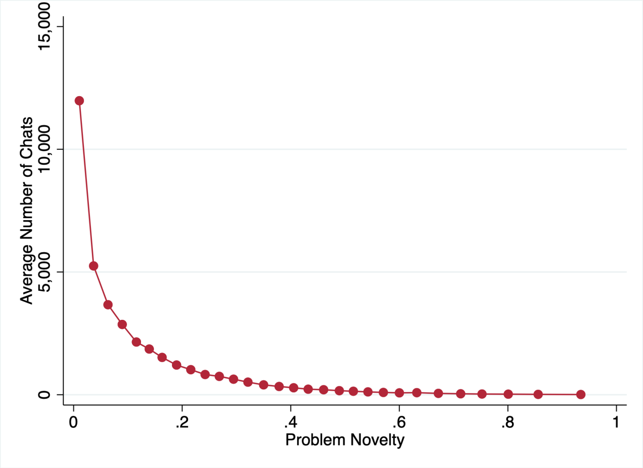


- **Agent tenure when ML starts:** agent experience when ML activated

Customers become nicer to international agents

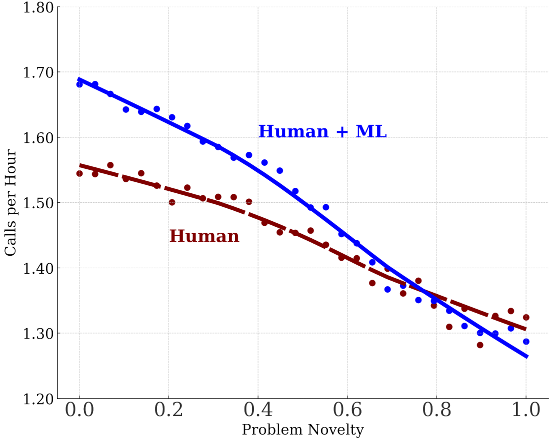


When is ML most useful?



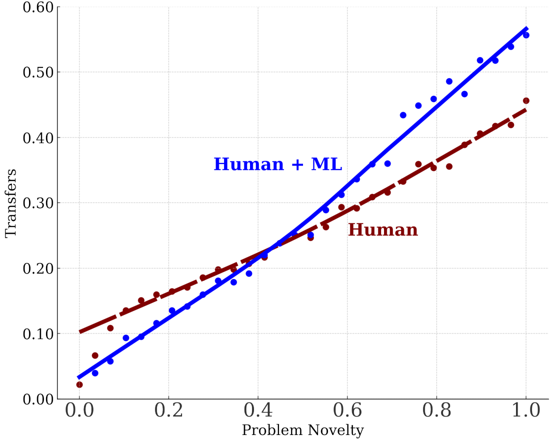
- Routine problems: "I forgot my password"
- Less common: "I need to add a new employee"
- One off: "Why was I flagged by my bank's risk management department?"

Return to ML is higher for more frequent problems



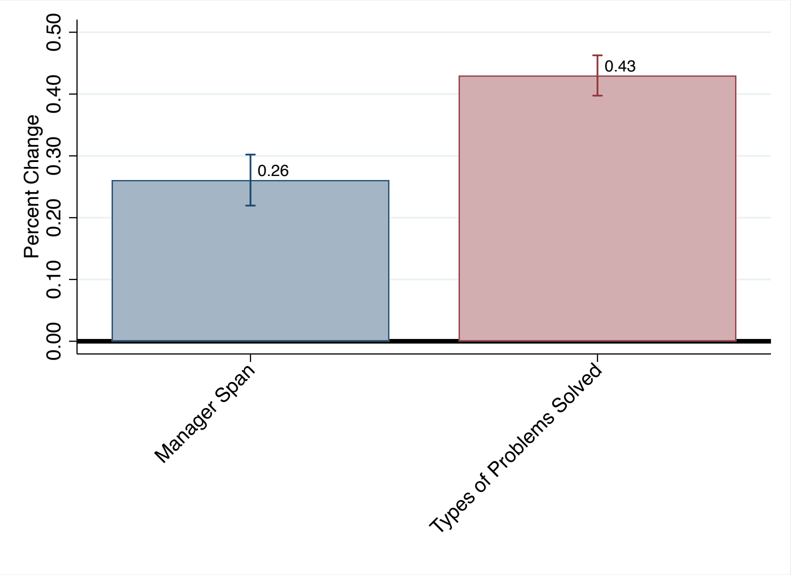
- On medium-common problems, human + ML outperform human alone
- On unusual problems from the tail of the distribution, human is faster

Spillover effects on requests for help



- Overall number of requests is flat
- On common problems, requests for help fall
- More bandwidth available for help on the uncommon problems
- Improving individual's decision making has spillover effects on those they ask for help (Athey et al. (1994), Garicano (2000))

Team size and communication patterns



Conclusion

AI has potential to change the nature and organization of work

1. Augmentation can increase productivity
2. Disproportionately driven ex-ante lower skill workers and outsourced workers
3. Accompanying organizational changes

Thank you!

- Questions? lraymond@mit.edu

Appendix

Assessing productivity improvements

Difference-in-difference specification

$$Y_{it} = \alpha_i + \gamma_t + \beta_1 (Treated_i x Post_{it}) + \epsilon_{it}$$

Stacked difference-in-difference specification

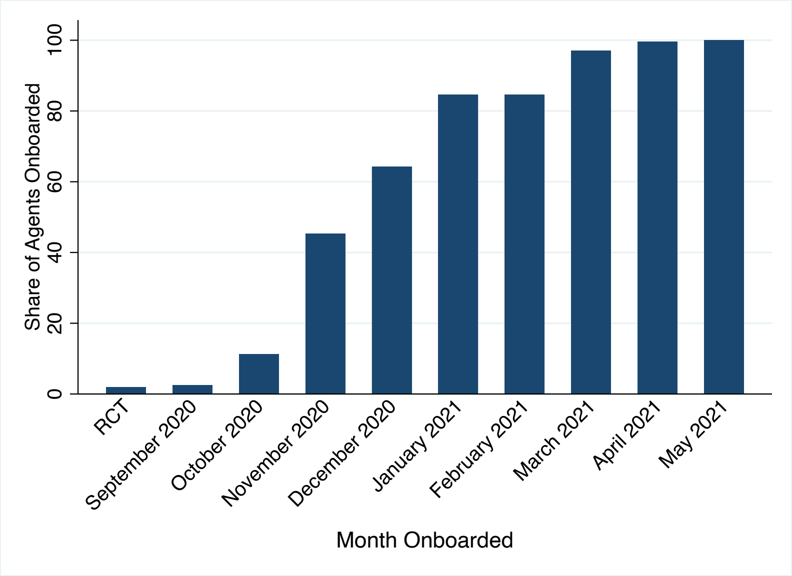
$$Y_{ite} = \alpha_{ie} + \sum_{\tau} D^{\tau}_{te} + \beta_1 (Treated_{ie} x Post^{\tau}_{ite}) + \epsilon_{ite}$$

- $Treated_i$ indicates agent i is ever treated with ML
- $Post_{it}$ is indicates when ML is turned on for agent i
- $Treated_{ie}$ indicates agent i is ever treated with ML in sub-experiment e
- $Post_{ite}$ is indicates when ML is turned on for agent i in sub-experiment e
- Stacked event study specification (Cengiz et al. (2019); Baker et al. (2021); Deshpande and Li (2019))

Assessing text changes

- **Embedding:** low dimensional vector representation of text that encodes meaning and complex characteristics of language
- Transformer model with attention mechanisms that weights words according to importance to better capture meaning
- **DistilBERT:** a smaller approximation of a transformer model (BERT)
- Enables similarity in language, meaning and sentiment analysis
- NLP model used in Google English language search

Deployment Timeline



Sample Summary Statistics

Variable	All	Control Agents	Treated Agents
Chats	3,758,698	374,731	2,635,864
Agents	6,846	1,035	1,813
Number of Teams	142	111	88
Share US Agents	.13	.095	.14
Distinct Locations	17	10	16
Average Chats per Month	158	112	212
Share Outsourced	.84	.62	.91
Number of Skills	2.8	2.3	3.3
Team Size	62	49	70
Average Call Duration (Min)	48	44	48
St. Dev. Call Duration (Min)	40	37	39
Issue Resolution Rate	79	77	81
Customer Satisfaction	62	62	61

Table: Sample Summary Statistics