July 2024

Productive Creators: A New Studio Indicator

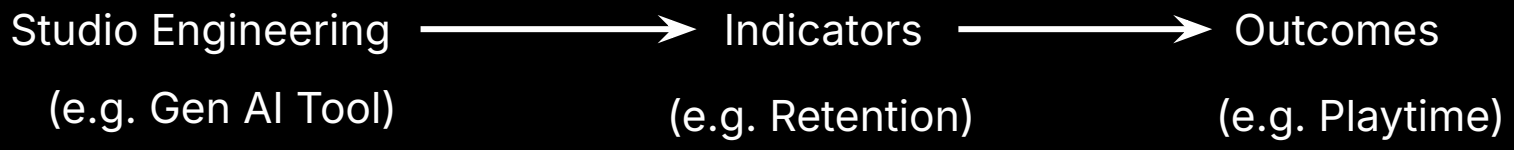
ROBLOX

Findings

- Productive Creator Metric = (Tenured + Engaged) x (Collaborating + Publishing)
  - 10% of MAU, 80% of Top/Core.
- It is superior to DAU, MAU, Time Spent.
  - Can explain 20% of changes in future playtime (agg).
  - Linked to 4 times the increase in playtime generated (user).
  - Linked to 3 times the increase in prob of breaking into top creator in future.
- It can guide Studio decisions:
  - Onboarding tutorial improved it.
  - Gen AI tools did not.
  - Reducing Login hangs on Mac did.

Takeaway: It matters how creators spent their time.

The Role of Indicators



What is a good indicator?

- Strong causal driver of business outcomes.
- Movable— connected to creator behaviours that can be nurtured.
- Predictable.
- Easy to build.
- Easy to communicate.

Studio Indicators

Indicator	Source Data	Size[1]	Definition	Concern
Total Creators	Time Spent	9.4M	Total users who downloaded Studio.	Too Broad
Active (L28)	Time Spent	3.7M	No. of users with time spent > 0.	Not quality adjusted.
Time Spent (L28)	Time Spent	392K Hours	Time spent by anyone with an active studio session.	Not quality adjusted.
New User Retention (L7)	Time Spent	60%	% active in current period, over those active in current & past.	Not quality adjusted.
Funnel (Playtime): Inactive, Active, Core, Top	Time Spent + Playtime	5.7M, 3.6M, 50K, 2.3K	Number of active creators with 0, 0-100, 100-100K, 100K+ amount of playtime in hours.	• Is actually an outcome • Not an intrinsic behaviour • Does not rule out luck
★ Productive Creators (L28)	Time Spent + Collaboration + Publishing +Tenure	0.42M	(Long Tenure OR Recently Engaged) AND (Collaborating OR Publishing)	Complex, involves multiple series and thresholds.

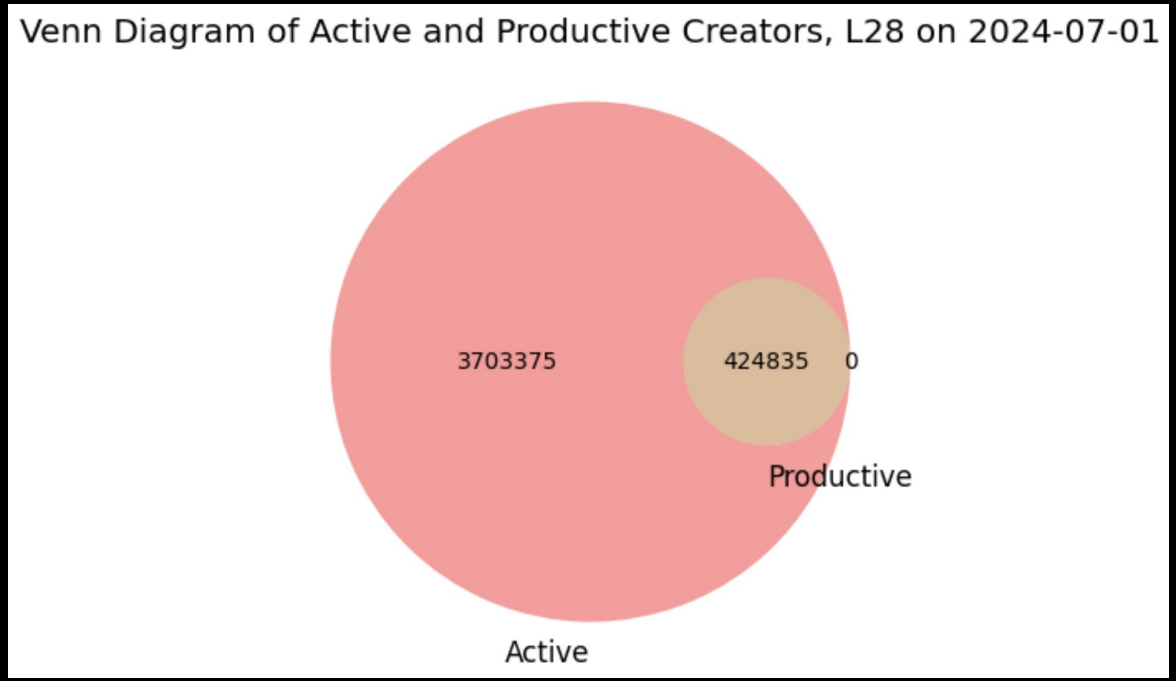
[1]Data is taken from dw_nonpii.developer_creator_velocity on ds = '2024-07-01'. Query link.

How many monthly active users are also productive?

Only ~10%.

Active:
studio_active_days_l28>0

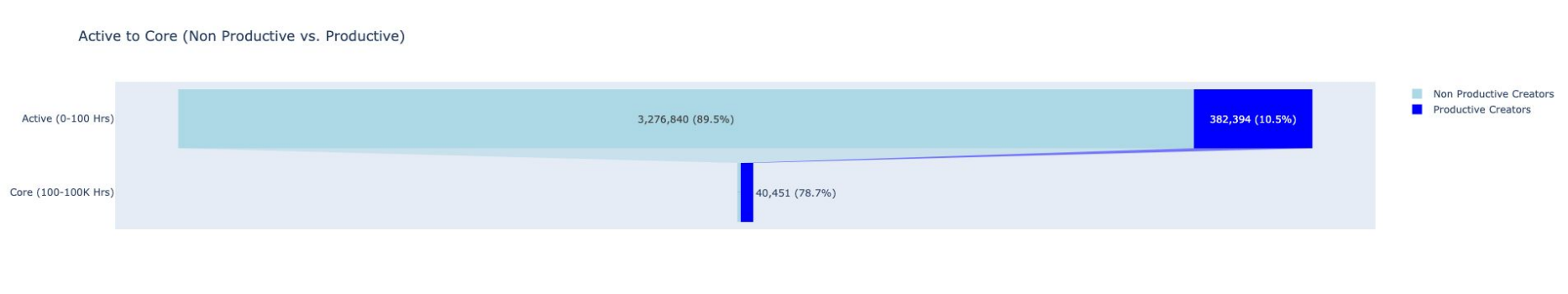
Productive:
studio_active_days_l28>4
AND
(collaboration_l28>1 OR
publish_l28>3)

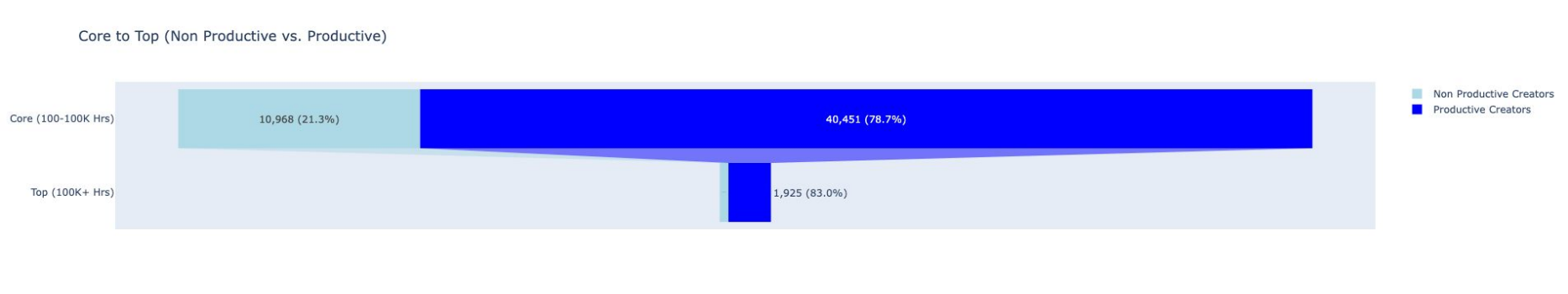


[1]Data is taken from dw_nonpii.developer_creator_velocity on ds = '2024-07-01'. Query link.

Funnel

About ~80% of current top/core creators are productive.





[1]Data is taken from dw_nonpii.developer_creator_velocity on ds = '2024-07-01'. Query link.

Studio Engineering → Productive Creators → Playtime

Platform Level Correlations

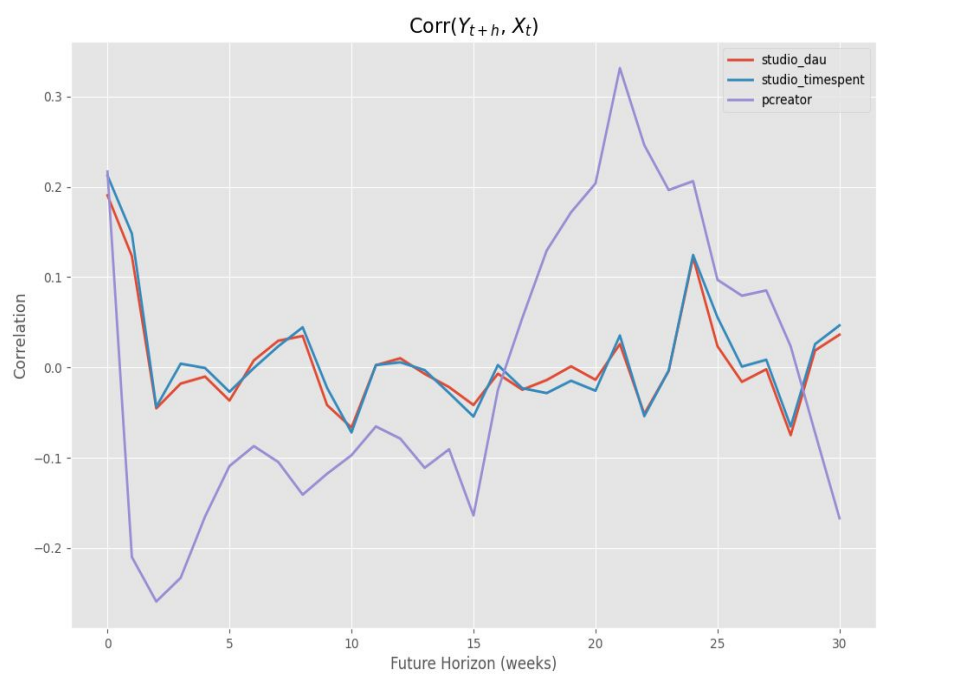
Y: Agg. Playtime Weekly Growth
X: Agg. Indicator Weekly Growth
Period: 2022-10 to 2024-07

<u>Productive Creators correlated to *future* playtime.</u>

Problem:
Confounding:

- X or Y (past) → X (today)
- X or Y (past) → Y (today)

Solution:
Control for past values.



Vector AutoRegression

Y: Agg Playtime WoW Growth
X: Agg Indicator WoW Growth

$$Y_t = \alpha^Y + \sum_r \beta_r^Y Y_{t-r} + \sum_r \tau_r^Y X_{t-r} + \epsilon_t^Y$$

$$X_t = \alpha^X + \sum_r \beta_r^X X_{t-r} + \sum_r \tau_r^Y Y_{t-r} + \epsilon_t^X$$

Notation:

- Alpha: Allows non-zero mean growth.
- Beta: Captures how “sluggishly” shocks pass through over time.
- Tau: Captures interdependence and delay in transmission.
- Epsilon: Capture driving forces unique to each series.
- Lag length “r”: determines how long do shocks persist in the system.

Tests of Predictability

- Granger Causality Test: Can series X predict *future* values of series Y?
  - Null: All Taus are 0.

$$Y_t = \alpha^Y + \sum_r \beta_r^Y Y_{t-r} + \sum_r \tau_r^Y X_{t-r} + \epsilon_t^Y$$

$$X_t = \alpha^X + \sum_r \beta_r^X X_{t-r} + \sum_r \tau_r^Y Y_{t-r} + \epsilon_t^X$$

Indicator	Indicator → Playtime (P-Value)	Playtime → Indicator (P-Value)
DAU	0.282	0.1350
Time Spent	0.177	0.1304
Productive	0.044	0.001

Only playtime and productivity predict each other.

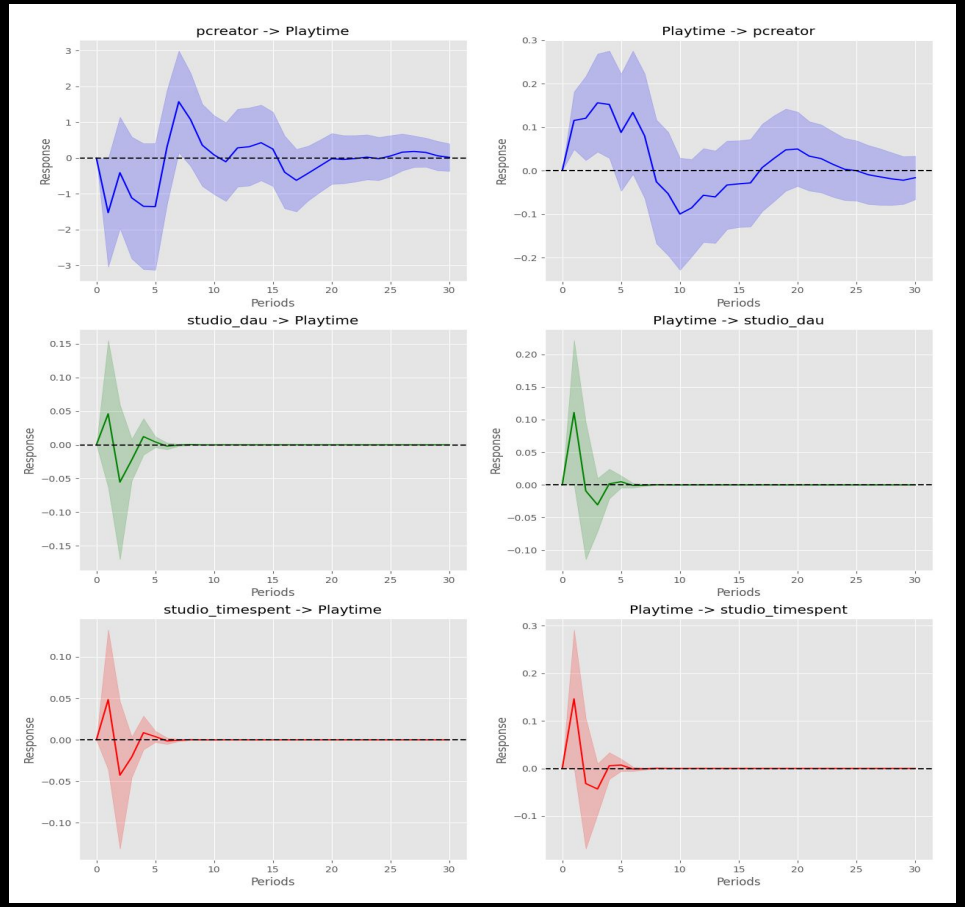
Full set of results: here.

Transmission Delays

IRF Plots: How shock to one series will pass through to other.

A 1 pp jump in productivity growth leads to a 1.5 pp jump in playtime growth in 8 weeks.

Indicator	Lag Order Chosen
DAU	2
Time Spent	1
Productive	9



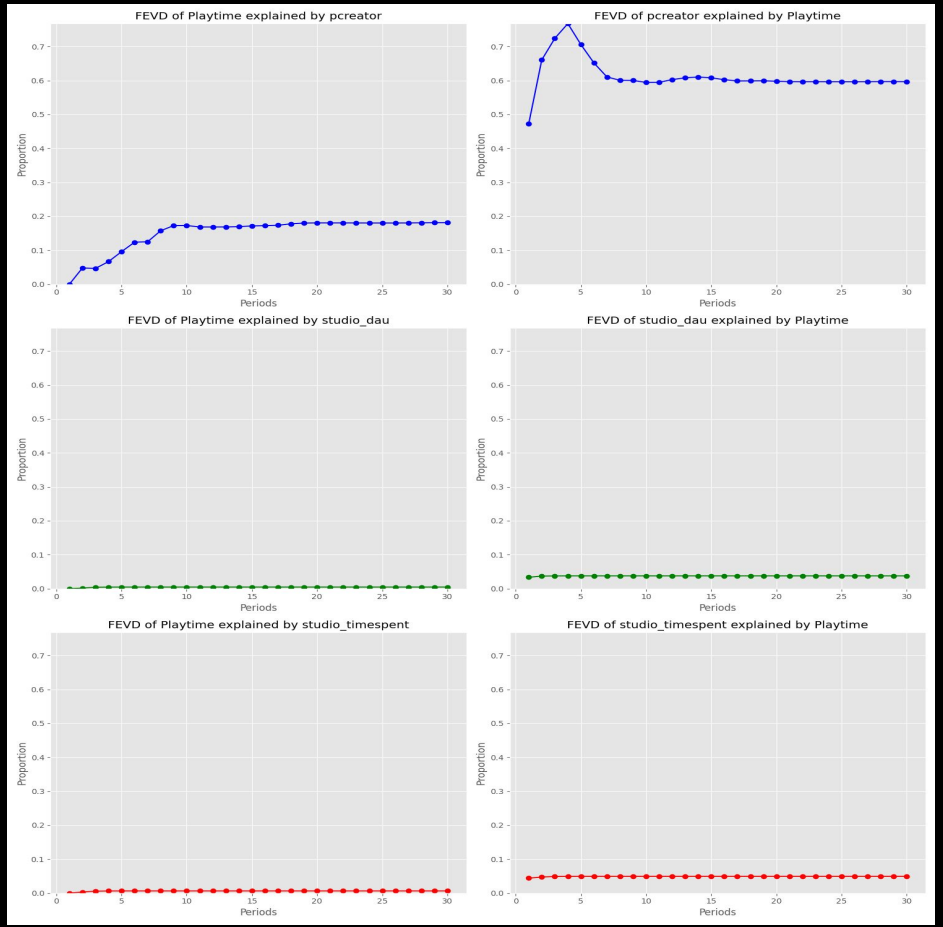
IRFs are obtained by simulating a single shocks to one series and then observing the system dynamics. Lag order is chosen by Akaike Information Criteria.

ROBLOX

Variance Explained

“Forecast Error Variance Decomposition” is how much one series is explained by shocks to the other series.

- Playtime can explain about 60% of the variation in future productivity.
- Productivity can explain about 20% of the variation in future playtime.
- Studio DAU and Time Spent are unrelated to Playtime.



ROBLOX Full set of results: here.

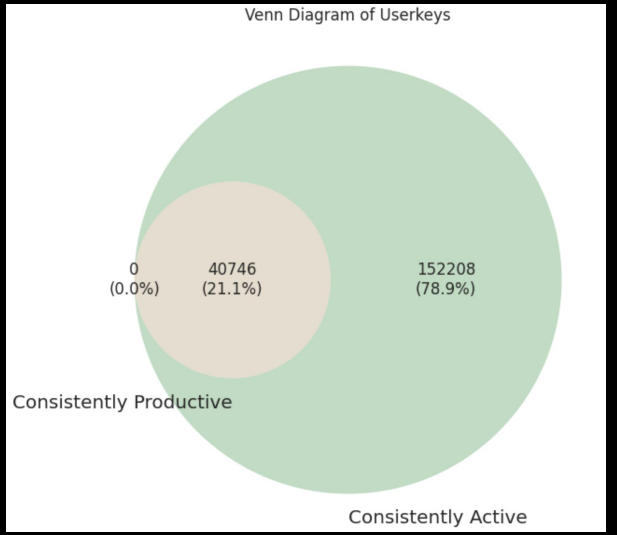
What if a creator was productive for a long time?

How much does it matter compared to mere activity?

Group 1 (Consistently Productive): Being a Productive Creator on the first date of every month, for 9 months.

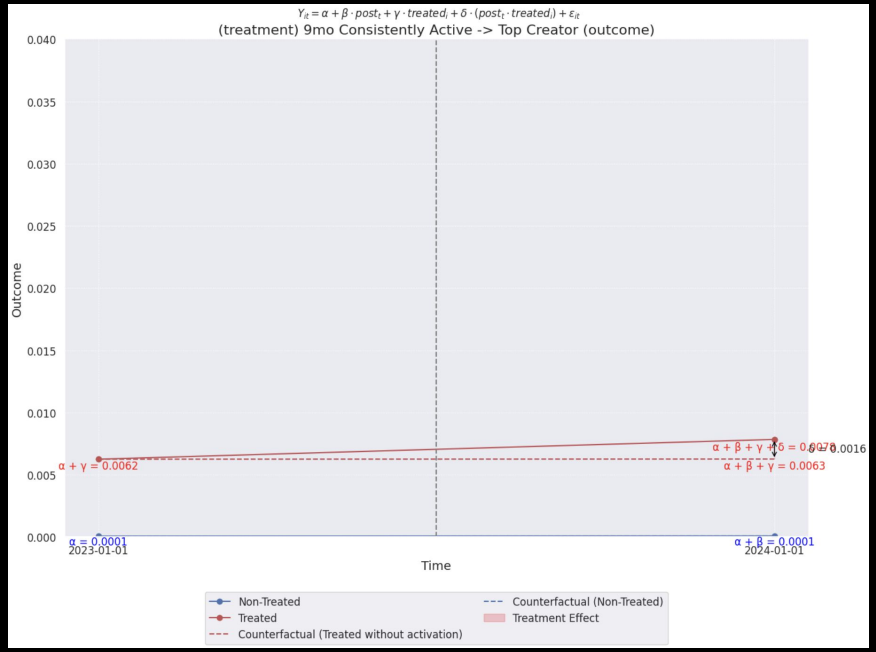
Group 2 (Consistently Active): Being a Active L28 at least once on first date of every month, for 9 months.

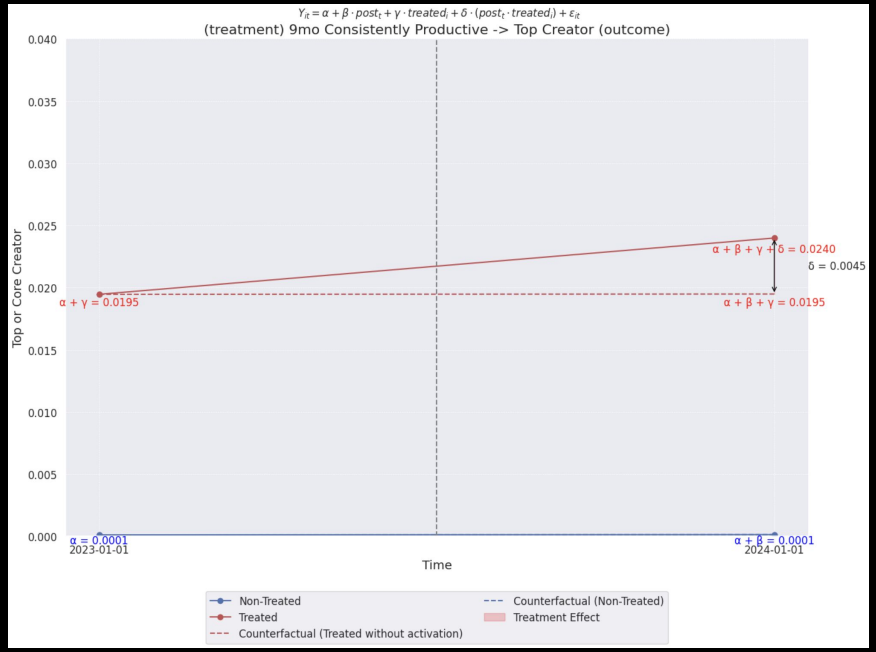
Period: 2023-02 to 2023-10



Active/Productive → Prob(Top Creator)

The impact of being “productive” on Prob(Top in Future) is ~280% more than simply being active.

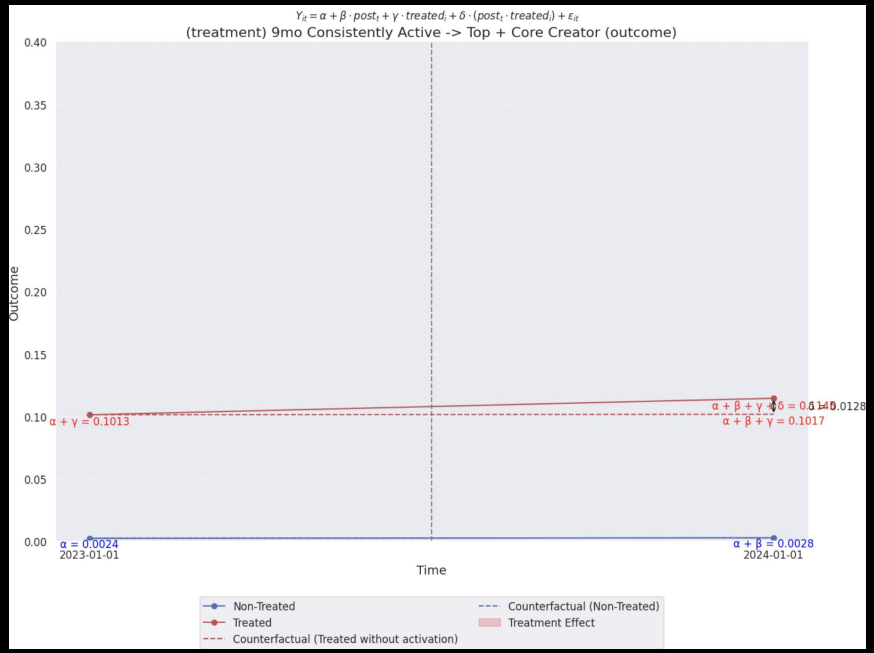


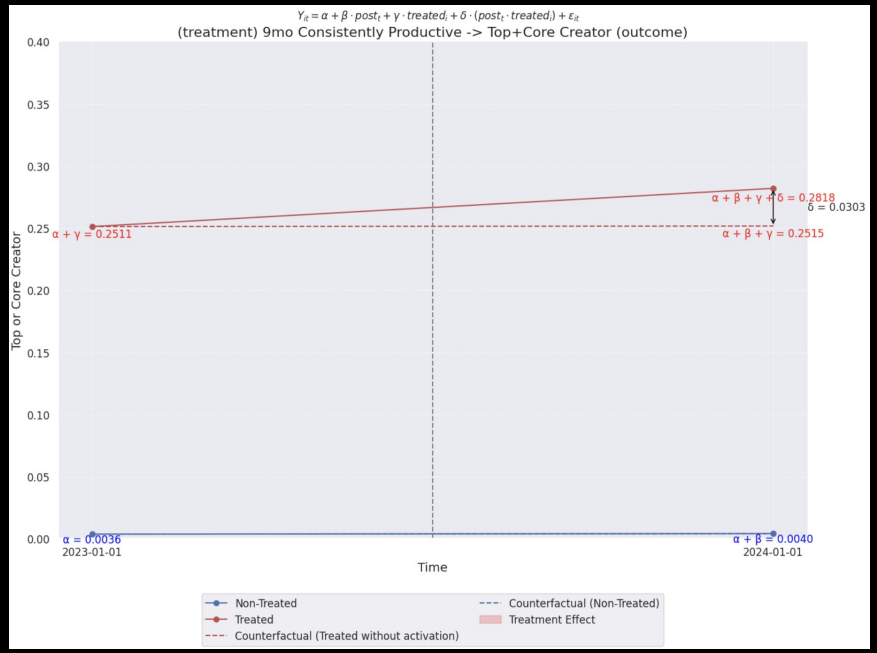


Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Active/Productive → Prob(Top or Core Creator)

The impact of being "productive" on Prob(Top/Core Future) is ~240% more than simply being active.

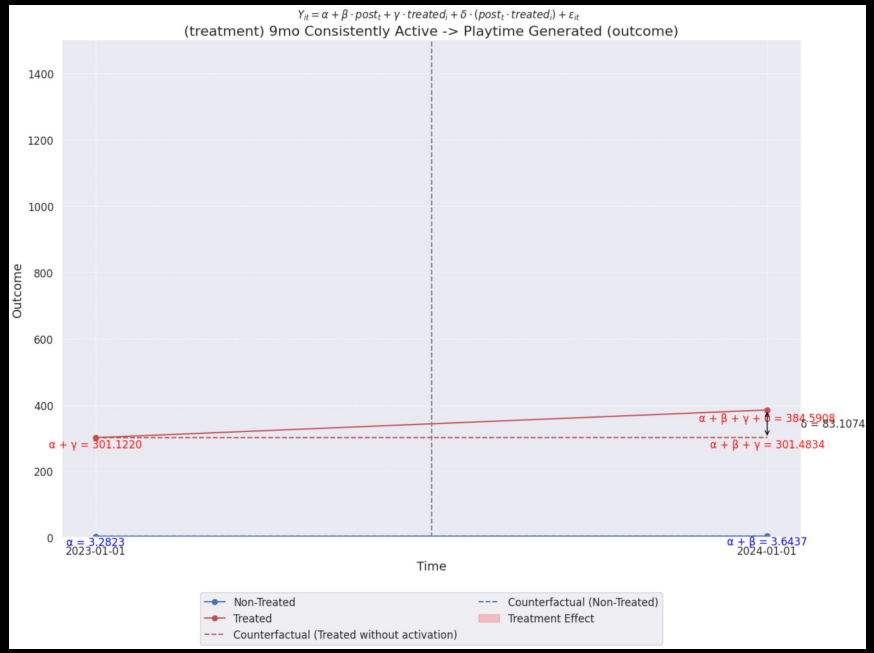


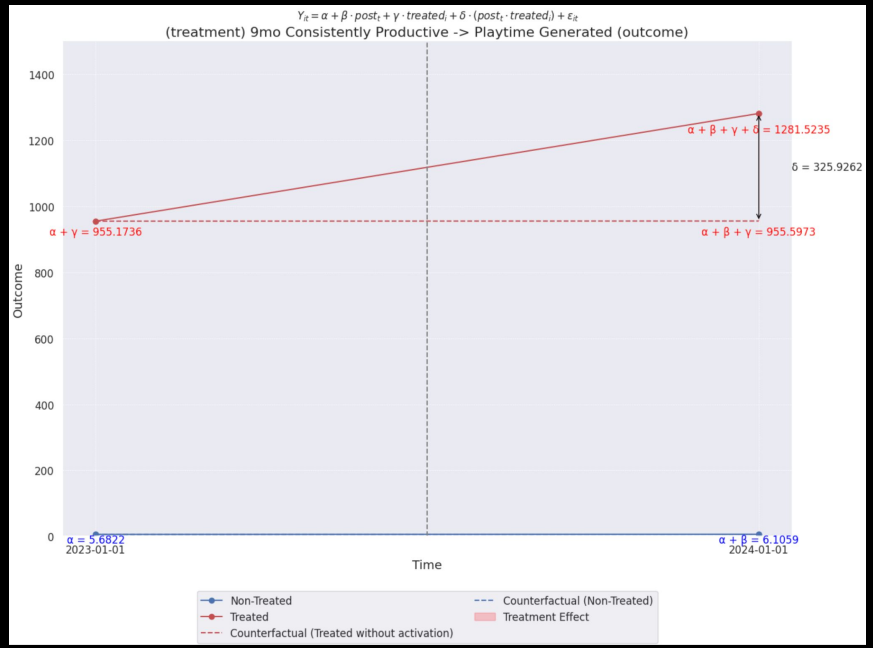


Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Active/Productive → Playtime Generated

The impact of being “productive” on Playtime Generated is ~400% more than simply being active.





Consistent productivity/activity captured at end of each month from 2023-02 to 2023-10. For outcomes, Pre Period: 2023-01 and Post Period: 2024-01.

Studio Engineering → Productive Creators → Playtime

Experiments

A/B Tests:

- Onboarding Tutorial: A new onboarding tutorial was only shown to 50% of new users
- Code Assist: Rollout of GPT or Llama based AI Autocomplete Assistance.

Natural Experiment:

- A series of bug fixes for Mac users in middle of May which did not extend to Windows.

Studio Onboarding Tutorial

Context: In 08-2023 a new Studio Onboarding Tutorial was launched as an A/B test aiming to improve retention in new studio users- about 50% users were impressed and 14% complied.

Eligibility: New Users who joined a month before the experiment.

Variables:

- Outcome Y: 1 if found Productive as of 9/1 (same period as experiment), else 0.
- Compliance T: 1 if user clicked on onboarding tutorial between 8/5 and 8/30, else 0.
- Enrollment Z: 1 if user was shown onboarding tutorial between 8/5 and 8/30, else 0.
- Covariates X: Pre-enrollment covariates, as of 8/4.

Studio Onboarding Tutorial: ATE and LATE

There is a pos stat sig effect of enrollment as well as compliance on productivity of creators. The impact of compliance is about 5 times as large.

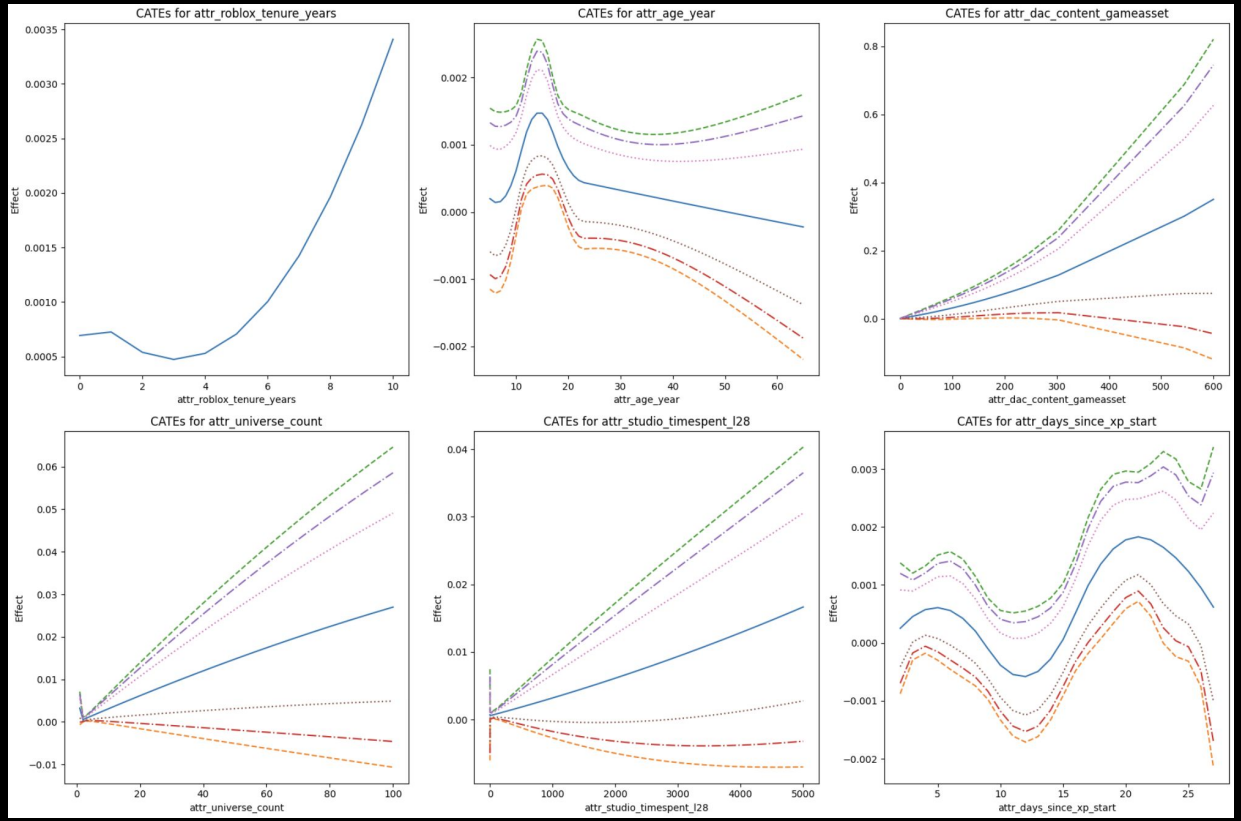
Estimand	Model	Equations	Estimate (b)	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	Y = a + b*Z + e	0.0007	0.0003	2.5536	0.0107
ATE for Enrollment	Multivariate OLS	Y = a + b*Z +c*D + e	0.0007	0.0003	2.4406	0.0147
ATE for Enrollment	DML	Y = b*Z +g(X) + e	0.0007	0.0003	2.386	0.017
ATE for Enrollment	DML (Interactive)	Y = g(Z,X) + e	0.0033	0.001	3.2118	0.0013
LATE for Compliance	Bivariate IV	Y = a + b*T + e T = c + d*Z + u	0.0025	0.001	2.5545	0.0106
LATE for Compliance	Multivariate IV	Y = a + b*T + c*X + e T = e + f*Z + g*X + u	0.0024	0.001	2.477	0.0132
LATE for Compliance	DML IV	Y = b*T + g(X) + e T = m(Z,X) + u	0.0023	0.001	2.3701	0.0178
LATE for Compliance	DML IV (Interactive)	Y = g(T,X) + e T = m(Z,X) + u	0.0007	0.0003	2.5536	0.0107

N=1.4 million, First stage: AUC of T on Z is 0.8, with X is 0.85.

Studio Onboarding Tutorial: CATE for Enrollment

Stat sig impact:

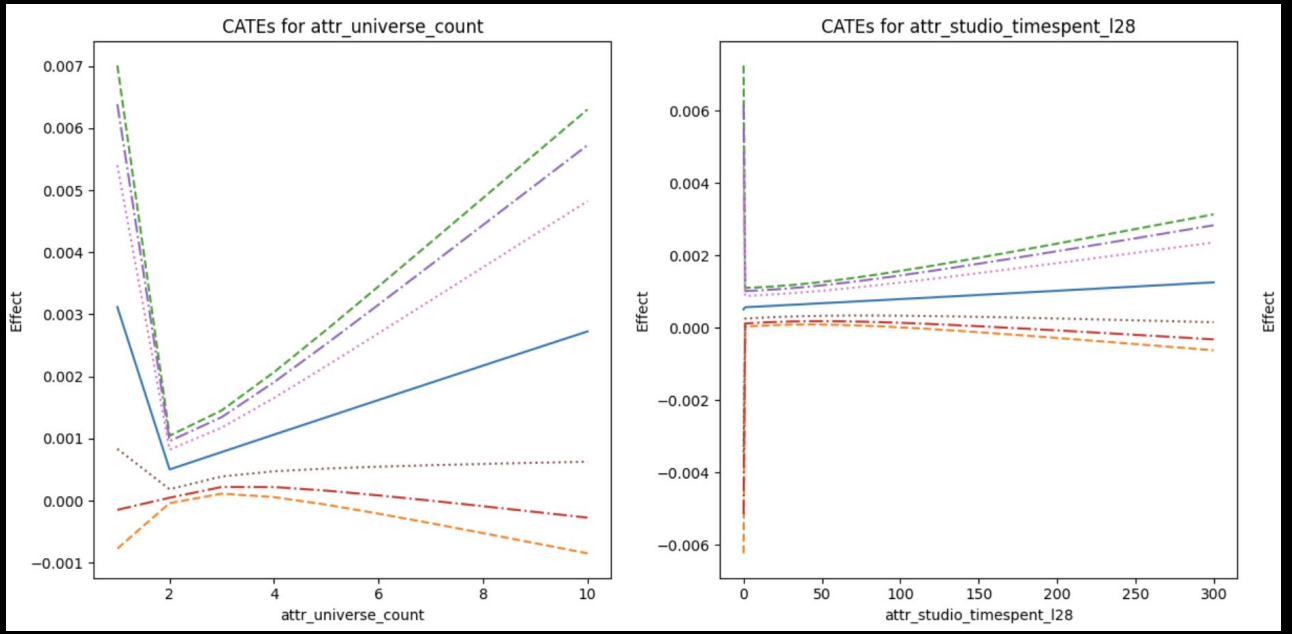
- Felt in 15 year olds
- Those who were enrolled later in the experiment. (confounded? Did those who are likely to be productive sign up later on)



Studio Onboarding Tutorial: CATE for Enrollment

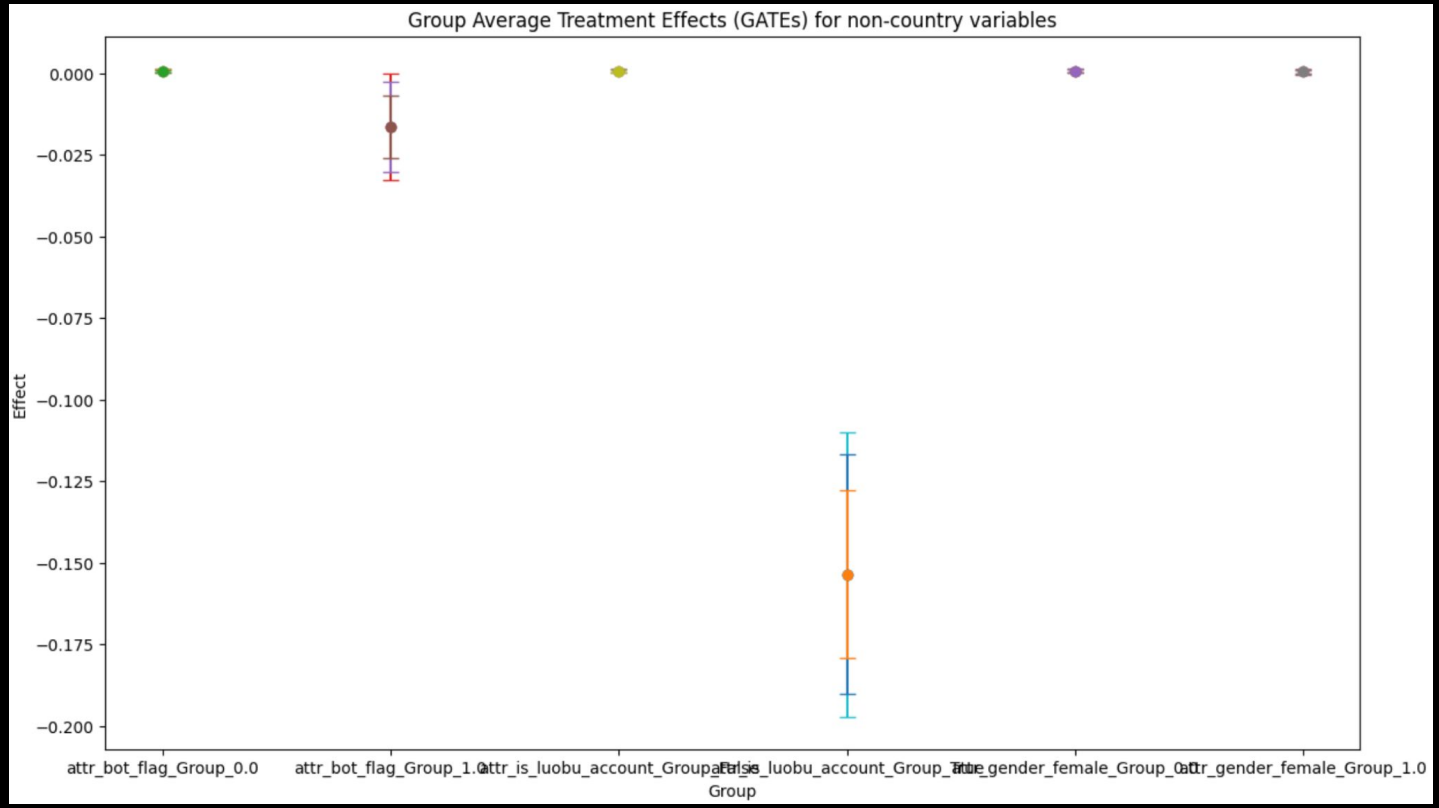
Pos stat sig impact:

- Those with 2-4 universes only.
- 0-50 seconds spent on Studio (completely new).

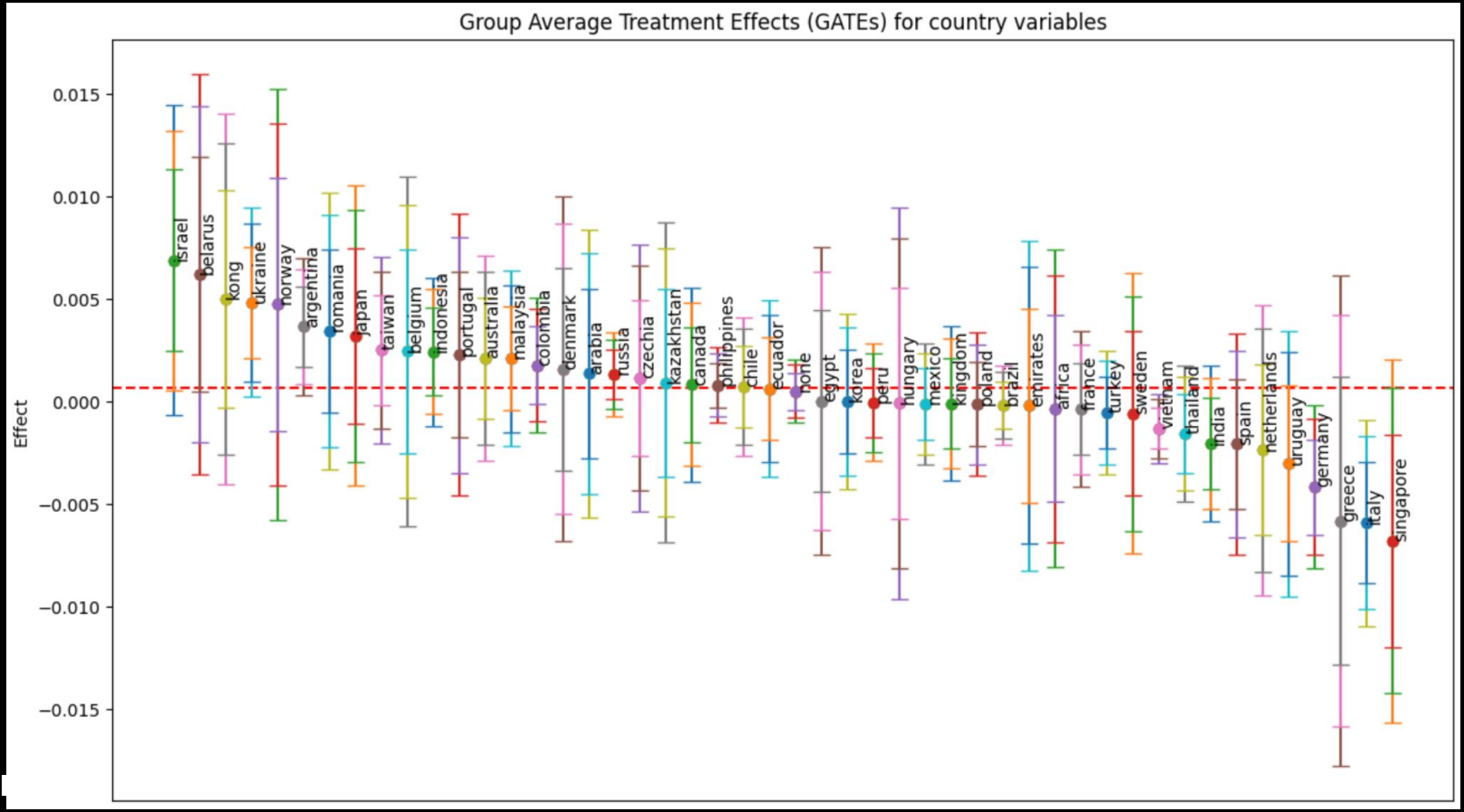


Studio Onboarding Tutorial: CATE for Enrollment

Bot and Luobo accounts registered a smaller impact than average.

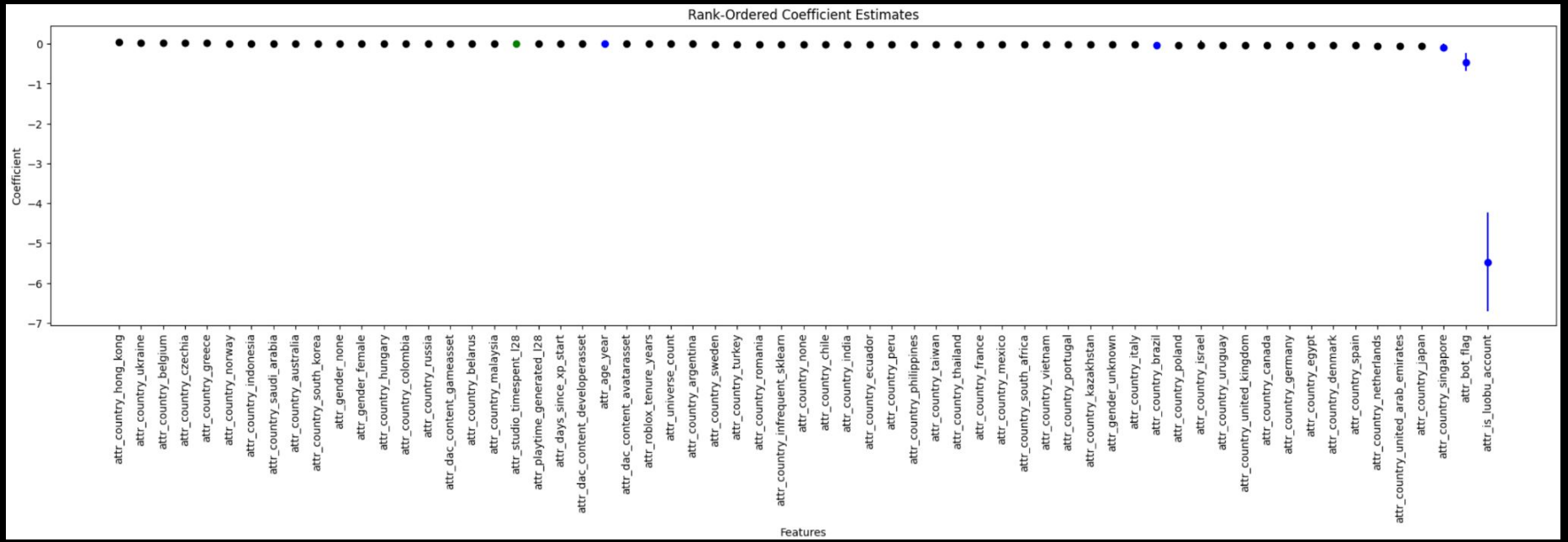


Studio Onboarding Tutorial: CATE for Enrollment



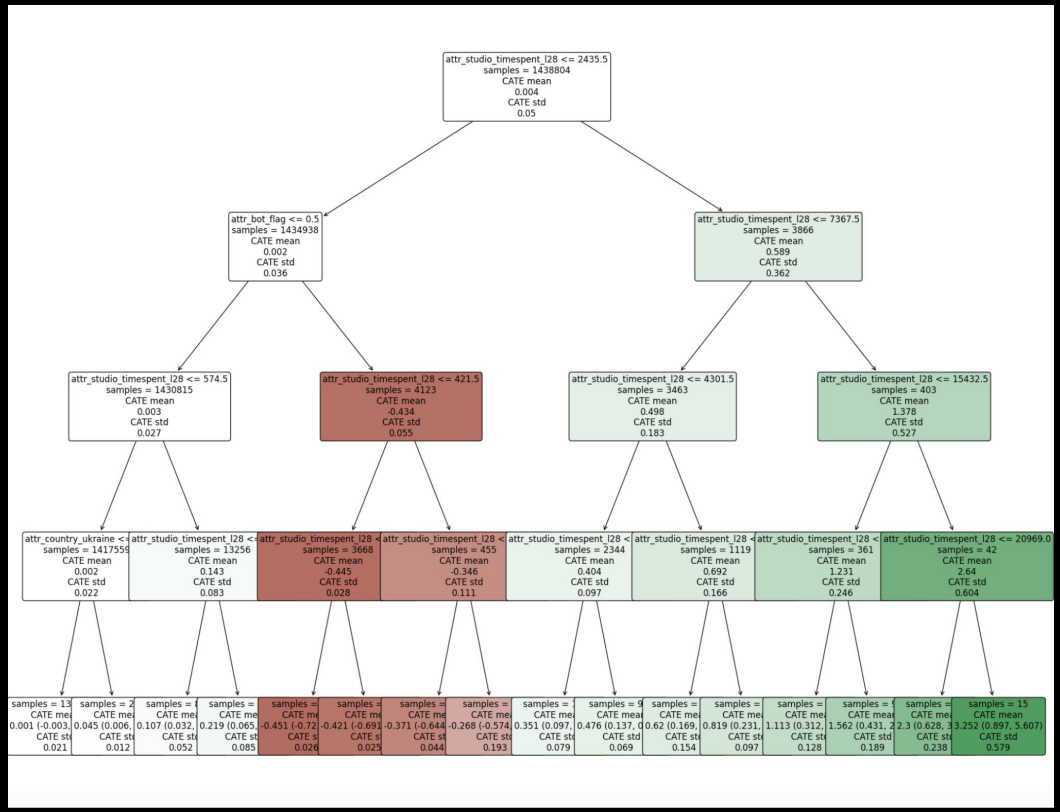
Studio Onboarding Tutorial: CATE for Compliance

- Linear Intent to Treat Double Robust Instrumental Variable Model
- Model: Y = \theta(X) T + g(X) + eps, \theta(X) is a linear function
- Those who already have some time spent or are younger stand to gain more.
- Some countries do better/worse compared to the US.



Studio Onboarding Tutorial: CATE for Compliance

- For non-bot accounts, the impact seems to be highest for those who have already spent 5 hours on the Studio.
- Bot users who have high time spent are likely to be less impacted by the tutorial.



Panel Evidence: Tutorial Activation → Productive Creator

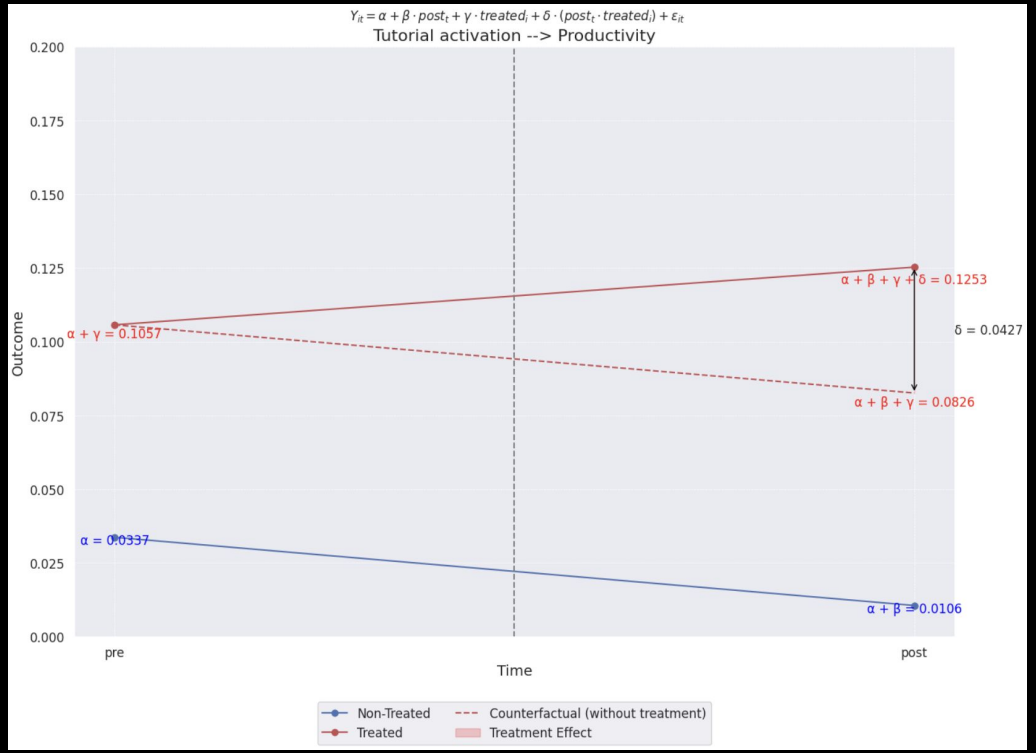
The impact of activating tutorial on productivity.

Treatment: Tutorial activation
Outcome: Productive Creator or Not
Data: as_of_dt from 23/06 to 23/12 (inclusive)

New users (base): month prior to as_of_dt
Pre period: month of as_of_dt
Post period: 3 months after as_of_dt
Treatment: Between pre and post months

Data:

A 2x2 Diff-in-Diff shows that activating tutorial is associated with a 4% pp increase in the probability of being productive.



Code Assist Experiment: Setup

In 2024, two versions of Code Autocomplete were rolled out to about 85k users.

Enrollment: When opening a script between 2/5 and 2/19.
- Control (20%): No Code Assistance
- Treatment 1 (40%): GPT-Based Assistance
- Treatment 2 (40%): CodeLlama (in-house) Assistance

Compliance:
- Activation: of code assistance, usually by trying to type something (96%).
- Completion: accepting recommendation from code assistance (90%).

Variables:
- Outcome Y: Productive Creator (as on 2/19)
- Enrollment Z: Enrollment into GPT/Codellama or None (between 2/5 and 2/19)
- Compliance T: Activation/Completion of Code Assistance (between 2/5 and 2/19)
- Covariates X: Pre-enrollment variables (as on 1/19).

Product page is here. Only have data on 75k out of 320k scripters due to majority not accepting terms of service.

First Stage Models

- Enrollment is indeed random but unable to predict outcome/compliance.
- Top covariates explaining outcome:
  - Studio time spent, past playtime generated, universe count, max contributors.
- Top covariates explaining compliance:
  - Studio time spent, age, country US, bot flag.

GPT

Model	Train AUC	Test AUC
Y on X, Z	0.791749	0.779711
Y on T, X	0.779286	0.776092
Y on X	0.800270	0.780080
Y on T	0.529282	0.538001
Y on Z, X	0.791749	0.779711
Y on Z	0.504667	0.494762
T on X, Z	0.729551	0.656748
T on X	0.692780	0.657436
T on Z	0.509247	0.493951
Z on X	0.589759	0.497097

CodeLlama

Model	Train AUC	Test AUC
Y on X, Z	0.774914	0.770914
Y on T, X	0.776347	0.772154
Y on X	0.795656	0.777867
Y on T	0.532480	0.531110
Y on Z, X	0.774129	0.769812
Y on Z	0.501084	0.494453
T on X, Z	0.731122	0.637216
T on X	0.701223	0.651515
T on Z	0.508654	0.495497
Z on X	0.549930	0.489151

GPT: ATE

We do not find stat sig effect of GPT code assist enrollment or compliance on productivity. Since bar for productivity is very high, we do note the positive direction.

Estimand	Model	Estimate	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	0.0058	0.0048	1.1902	0.234
ATE for Enrollment	Multivariate OLS	0.0047	0.0045	1.0403	0.2982
ATE for Enrollment	DML	0.0056	0.0042	1.3322	0.1828
ATE for Enrollment	DML (Interactive)	0.0063	0.0043	1.4663	0.1426
ATE for Compliance	Bivariate IV	1.1459	1.1216	1.0217	0.3069
ATE for Compliance	Multivariate IV	1.004	1.1216	0.8952	0.3707
ATE for Compliance	DML IV	1.4453	1.4521	0.9953	0.3196
ATE for Compliance	DML IV (Interactive)	1.3256	1.2582	1.0536	0.2921

CodeLlama: ATE

Not even a directional impact. This raises a conjecture that GPT may be superior to Code Llama w.r.t creator productivity.

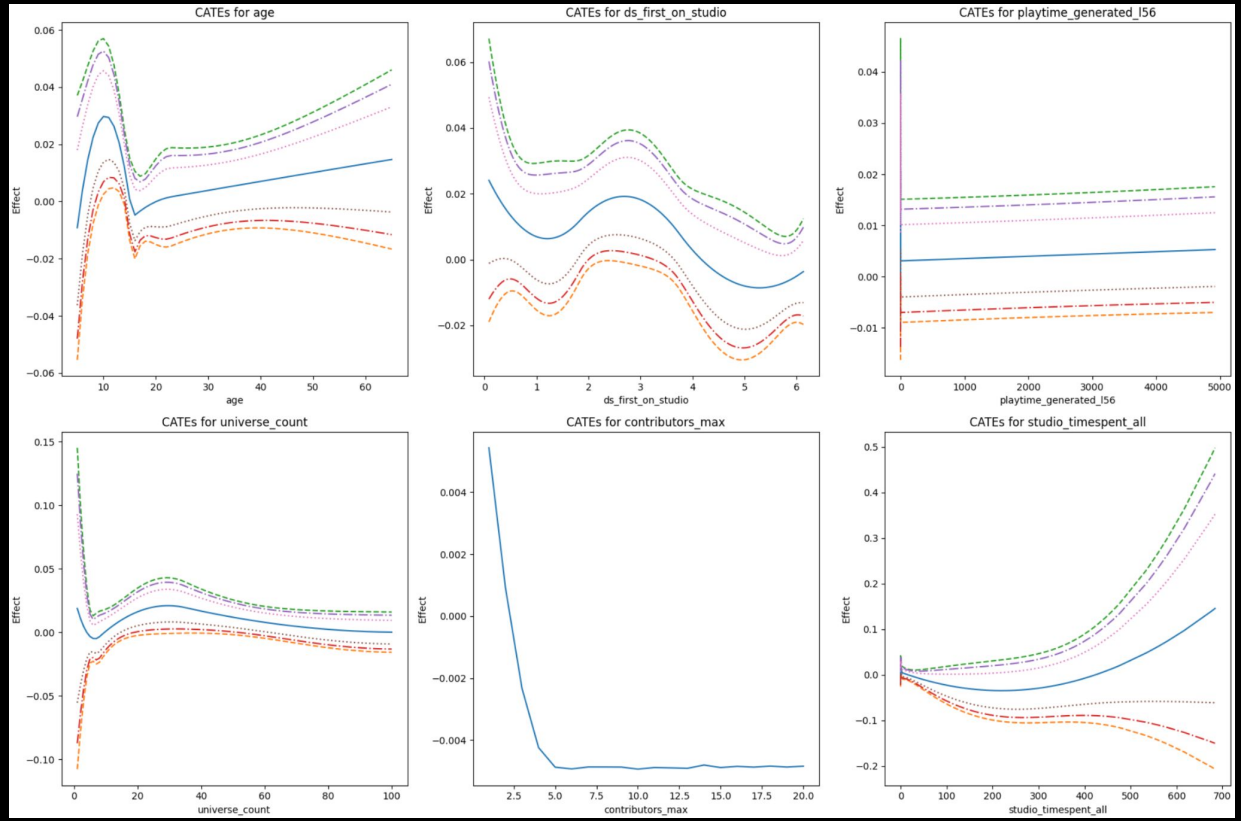
Estimand	Model	Estimate	Std. Error	t-value	p-value
ATE for Enrollment	Bivariate OLS	0.0005	0.0048	0.104	0.9172
ATE for Enrollment	Multivariate OLS	0.0000	0.0045	-0.0087	0.993
ATE for Enrollment	DML	-0.0003	0.0042	-0.0771	0.9386
ATE for Enrollment	DML (Interactive)	-0.0004	0.0043	-0.0924	0.9264
ATE for Compliance	Bivariate IV	0.1019	0.976	0.1044	0.9168
ATE for Compliance	Multivariate IV	-0.0085	0.9732	-0.0087	0.993
ATE for Compliance	DML IV	-0.0605	0.9155	-0.066	0.9473
ATE for Compliance	DML IV (Interactive)	-0.0667	0.9183	-0.0727	0.9421

GPT: CATE

Stat sig impact:

- Felt in 10 year olds
- Those with roughly 3 years of tenure.

Pos directional impact in those with few universes, shorter tenure.

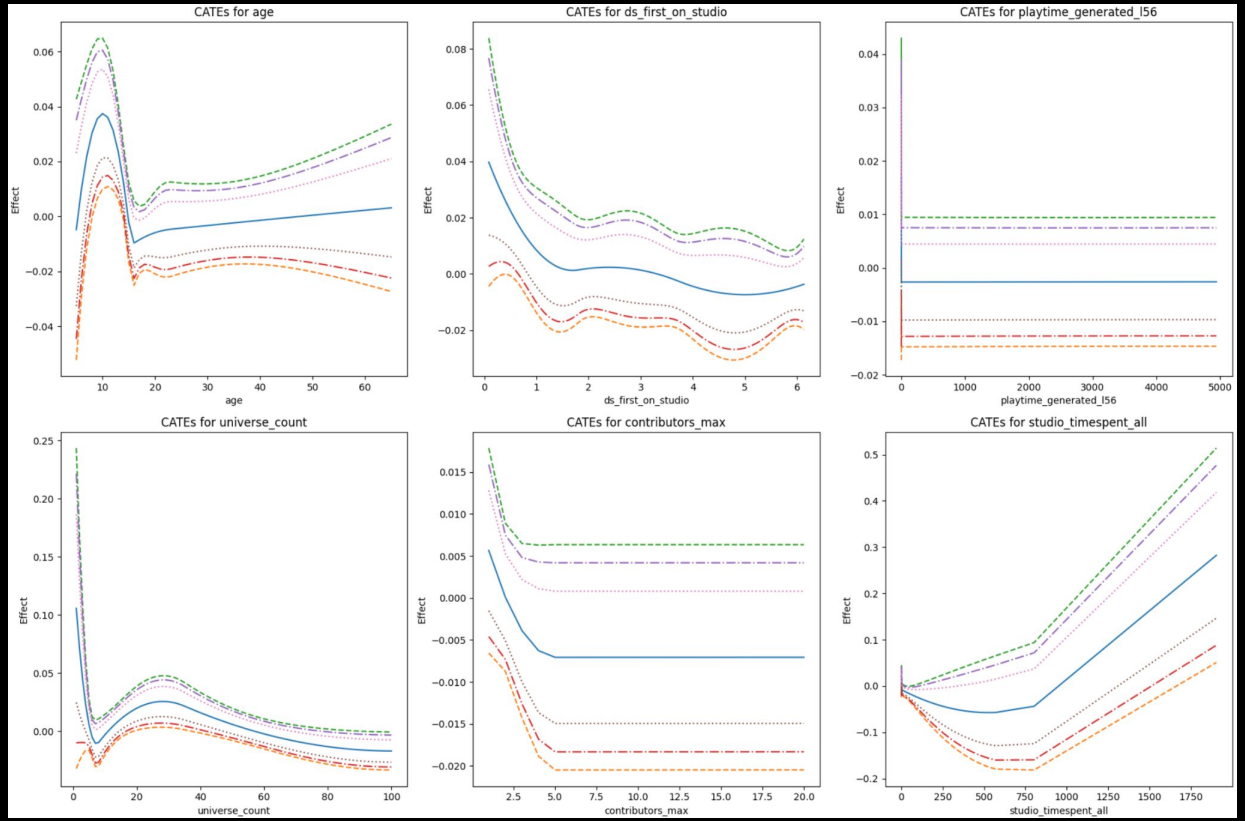


CodeLlama: CATE

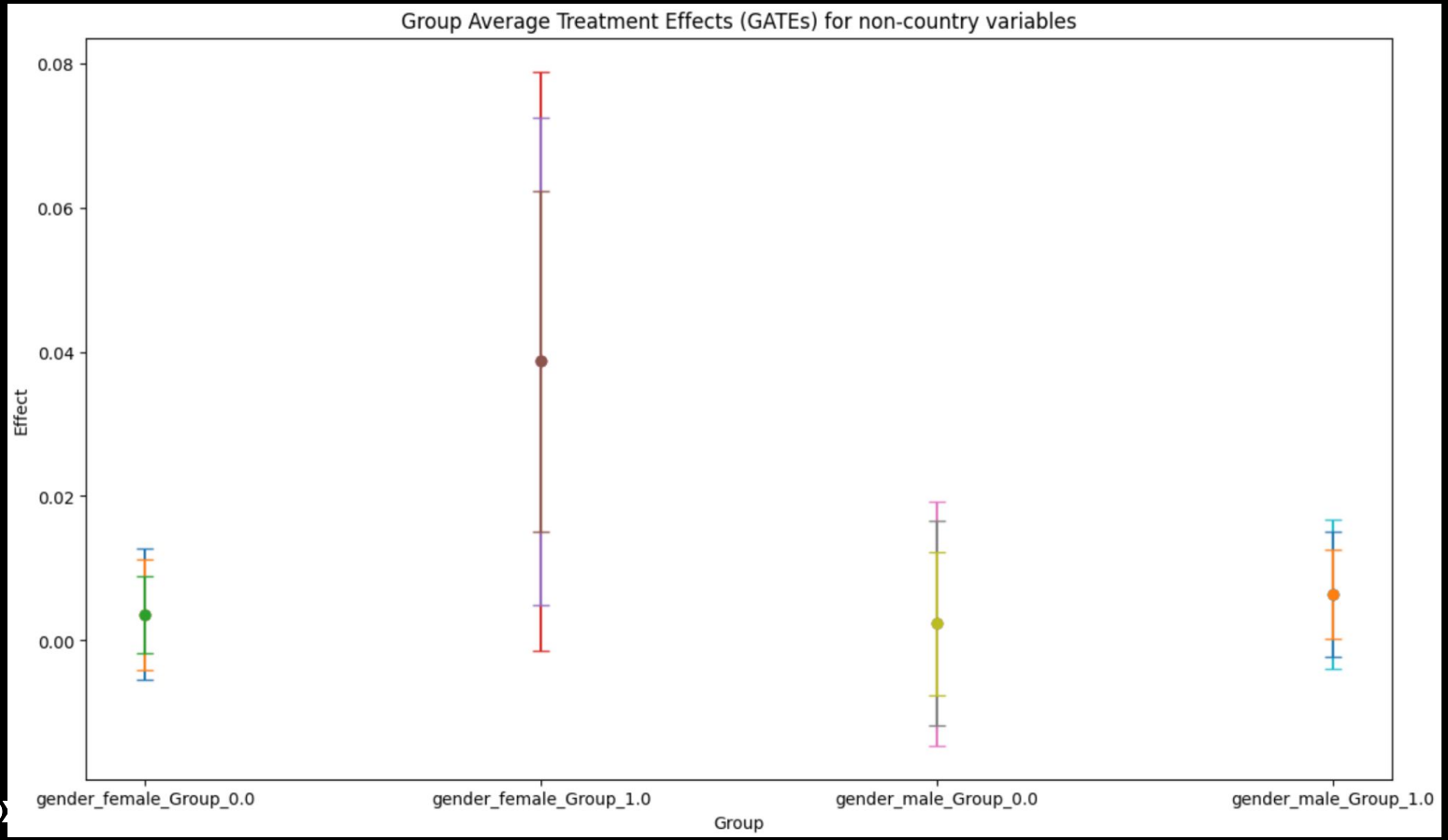
Stat sig impact:

- Felt in 10 year olds
- Those with >5 hours of studio time spent.
- Those with >20 but <60 universes.

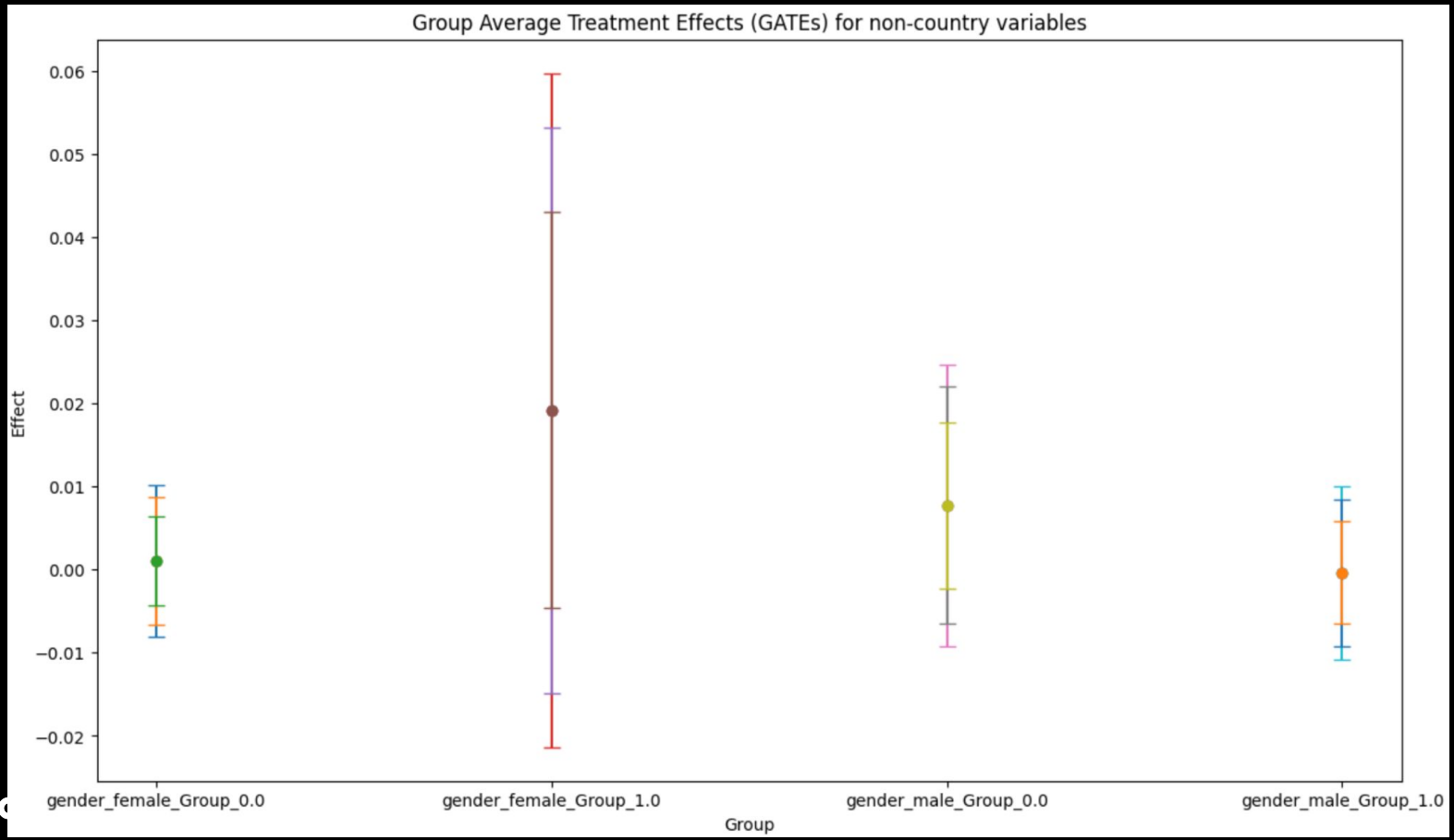
Pos directional impact in those with few universes, shorter tenure.



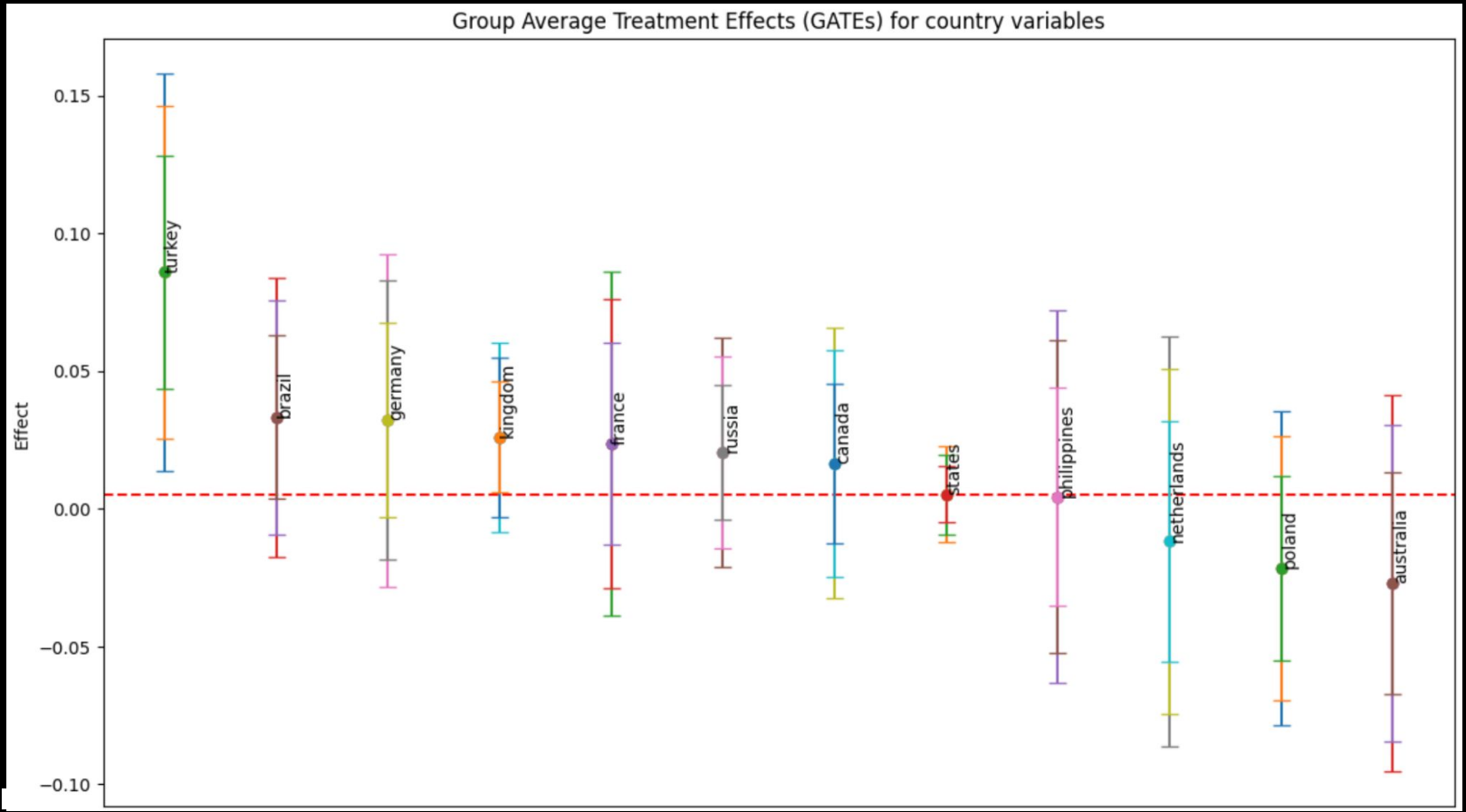
GPT: Gender


Group Average Treatment Effects (GATEs) for non-country variables
0.08
0.06
0.04
0.02
0.00
Effect
gender_female_Group_0.0
gender_female_Group_1.0
gender_male_Group_0.0
gender_male_Group_1.0
Group

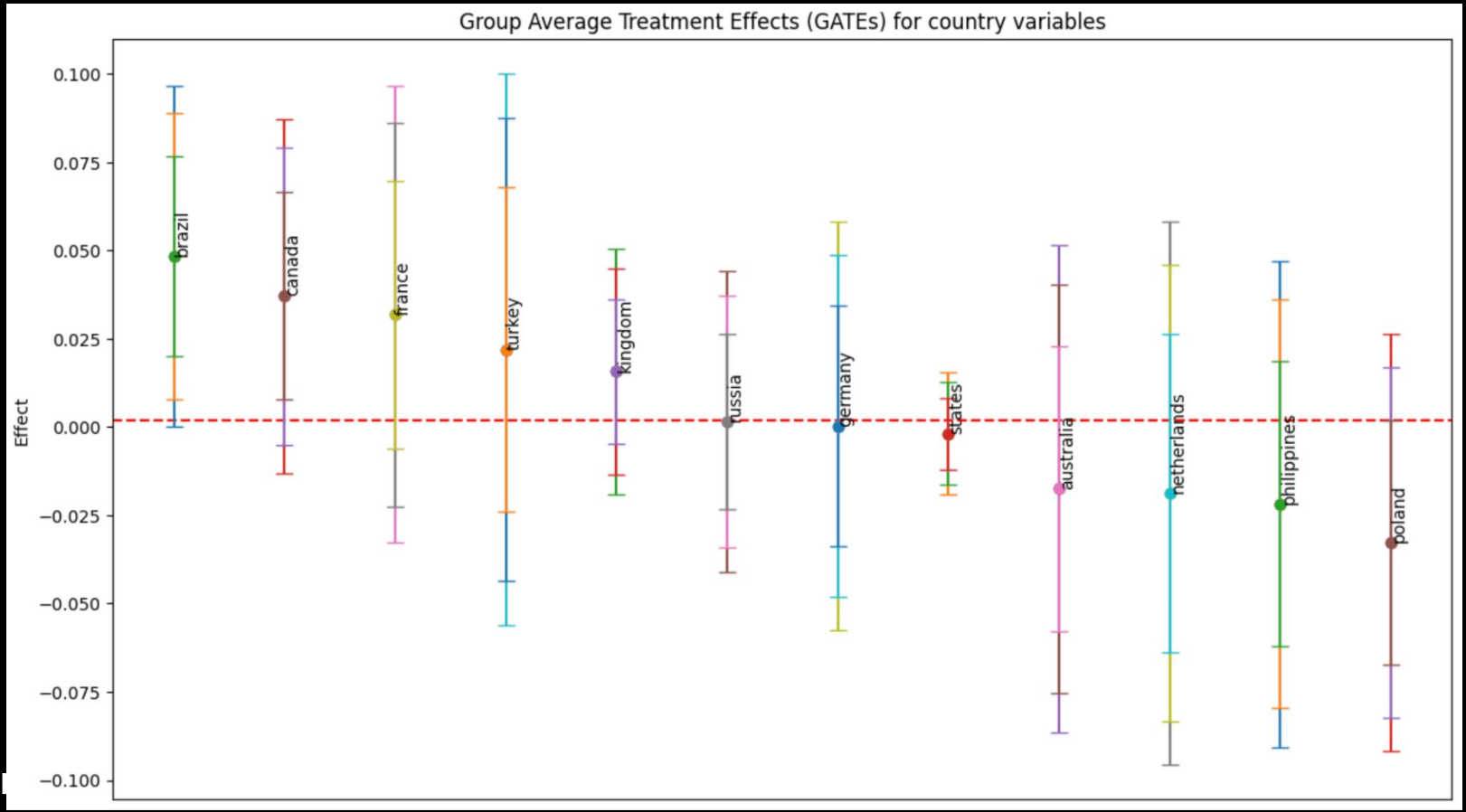
CodeLlama: Gender



GPT: Countries

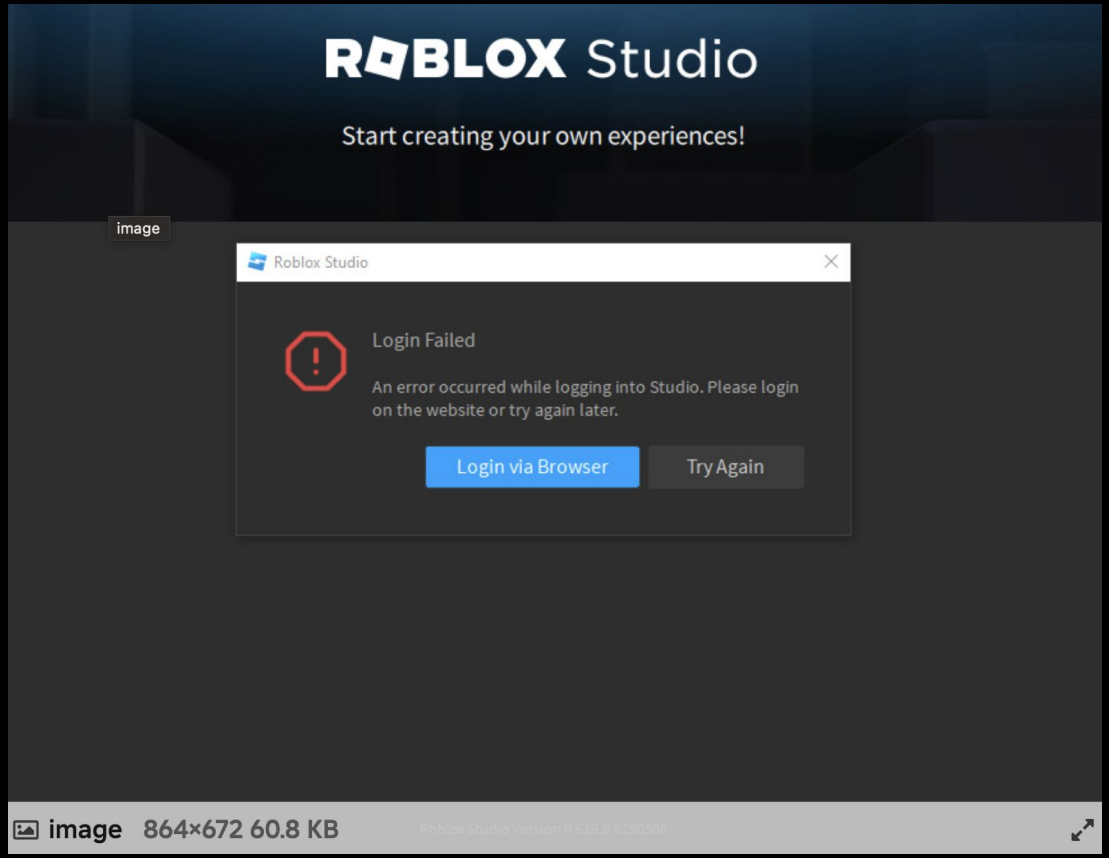


CodeLlama: Countries



Mac Bug Fixing

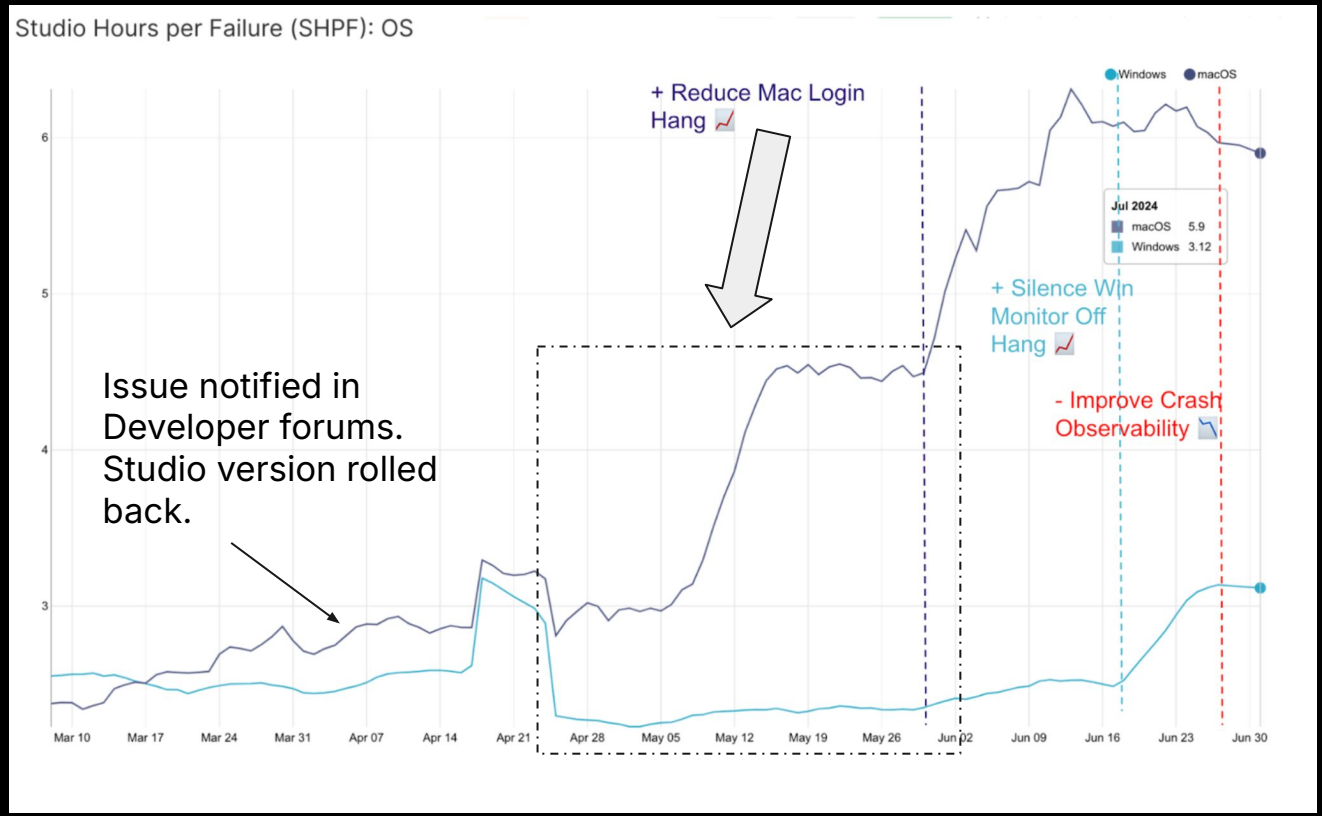
In April large number of Mac users began to face serious issues at login.



Mac Bug Fixing: Studio Hours Per Failure

Focused bug fixing in mid-May led to approx 30% inc in the SHPF for Mac users.

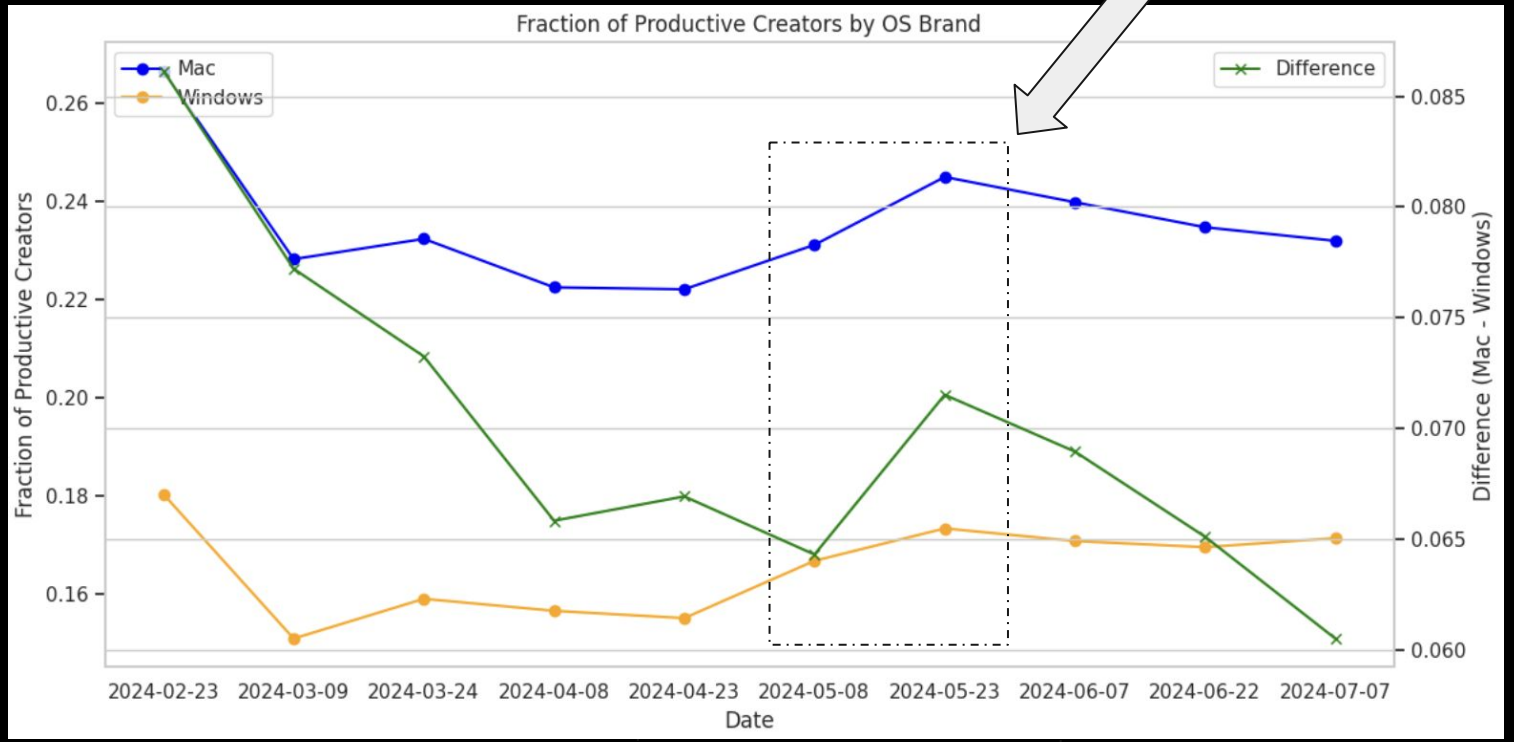
Context: Engineers confirmed that no major update went out to Windows in May, and that this fix would majorly benefit Mac users, esp. who had not logged in for a long time.
[Dev forum link.]
[Git commit link.]



ROBLOX

Mac Bug Fixing: Productivity (L12) with L3 Eligibility

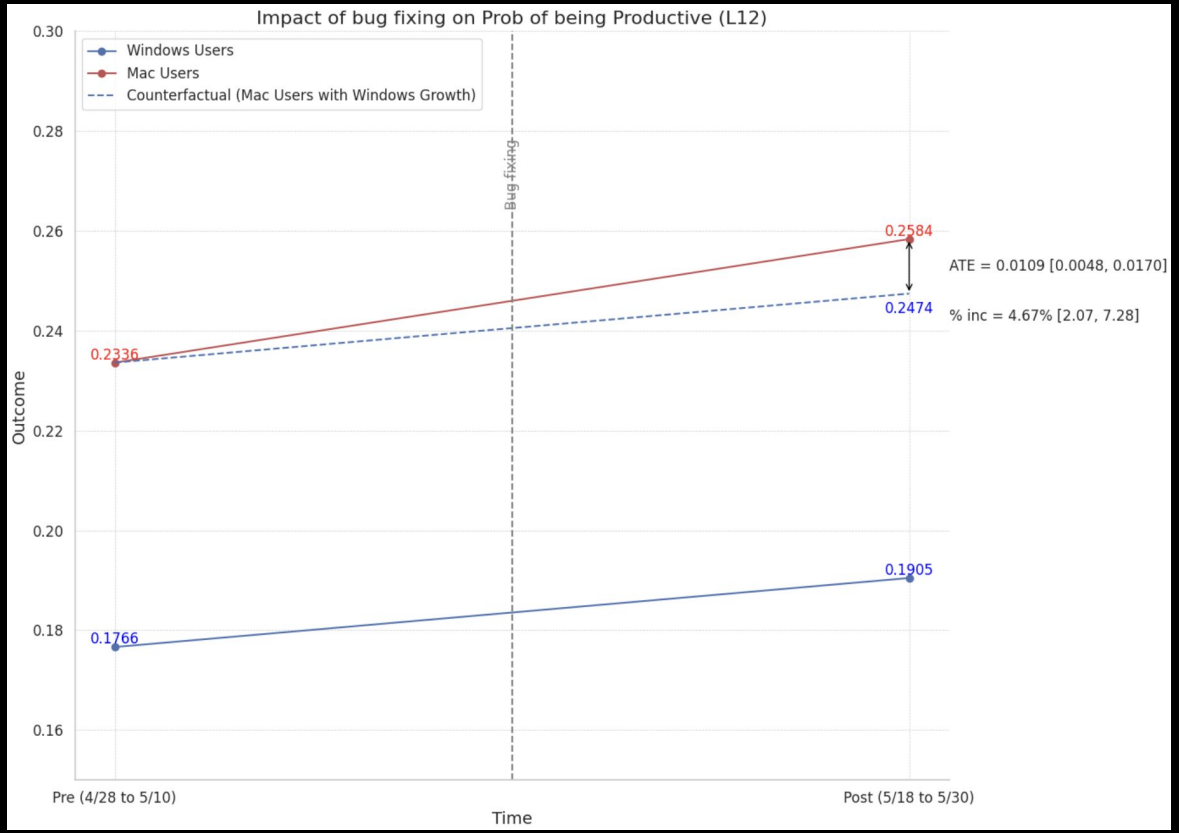
Secular decline in the difference. Thus the estimate will actually be conservative.



Mac Bug Fixing: IV + Diff-in-Diff

A 5% stat sig increase in productivity in mac users, due to bug-fixing.

Thus a 10% increase in SHPF is associated with a 2% increase in Productivity in Mac users.



Eligibility: Logged into Studio 3 days prior to the pre/post windows.

ROBLOX

Road ahead

- Decomposing collab/publishing by actions (script, build, design, test, etc.)
- Optimal thresholds for “productivity” definition
- Measuring creator conscientiousness - code quality, efficiency, consistency, etc.
- Study how productivity translates into playtime across genres.
- Linking productivity to creator milestones.