Position: AI/ML Influencers Have a Place in the Academic Process

Iain Xie Weissburg*, Mehir Arora*, Xinyi Wang, Liangming Pan, William Yang Wang

Data Collection

We collect over **8,000 unique papers** shared by **2 major influencers** over **5 years**.

Our influencers **share >95% of papers before any citations** and **all shared papers are open access**.

We build an unbiased control set for each influencer by **matching on 9 explicit covariates**.

Matching Covariates

```
year, venue, open_access, n_authors,
first(h-index), max(h-index),
first(citations), max(citations),
topic.embedding
```

Number of Treatment Papers

	AK	Komatsuzaki
# Unique Papers	9,171	1,273
# with all attributes	8,259 (90%)	1,191 (94%)
# Matched Papers	6,890 (75%)	955 (78%)

Matched Pair Review Scores
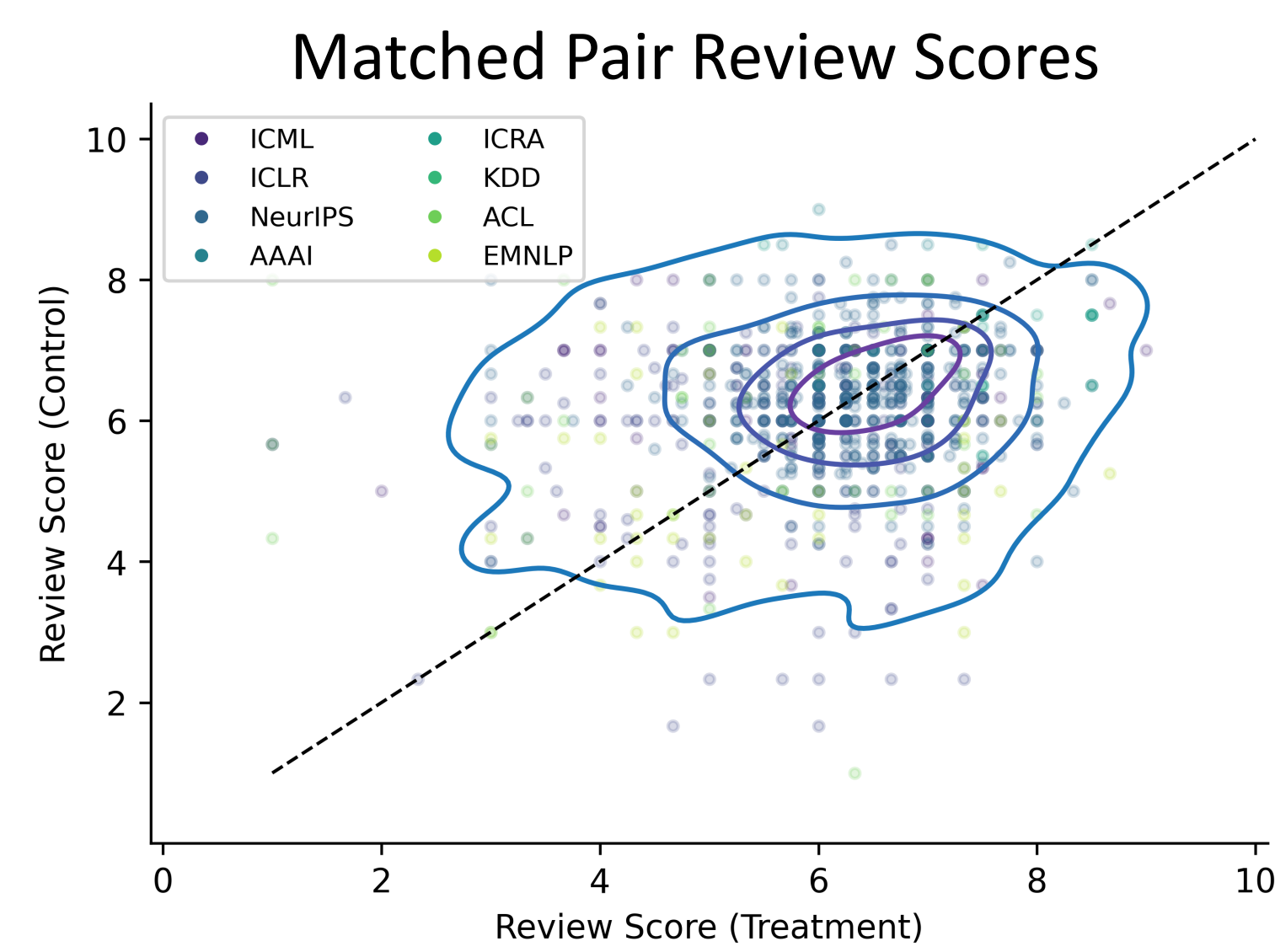
The subset of our dataset with OpenReview scores is **not significantly different from the overall dataset** (p > 0.05).

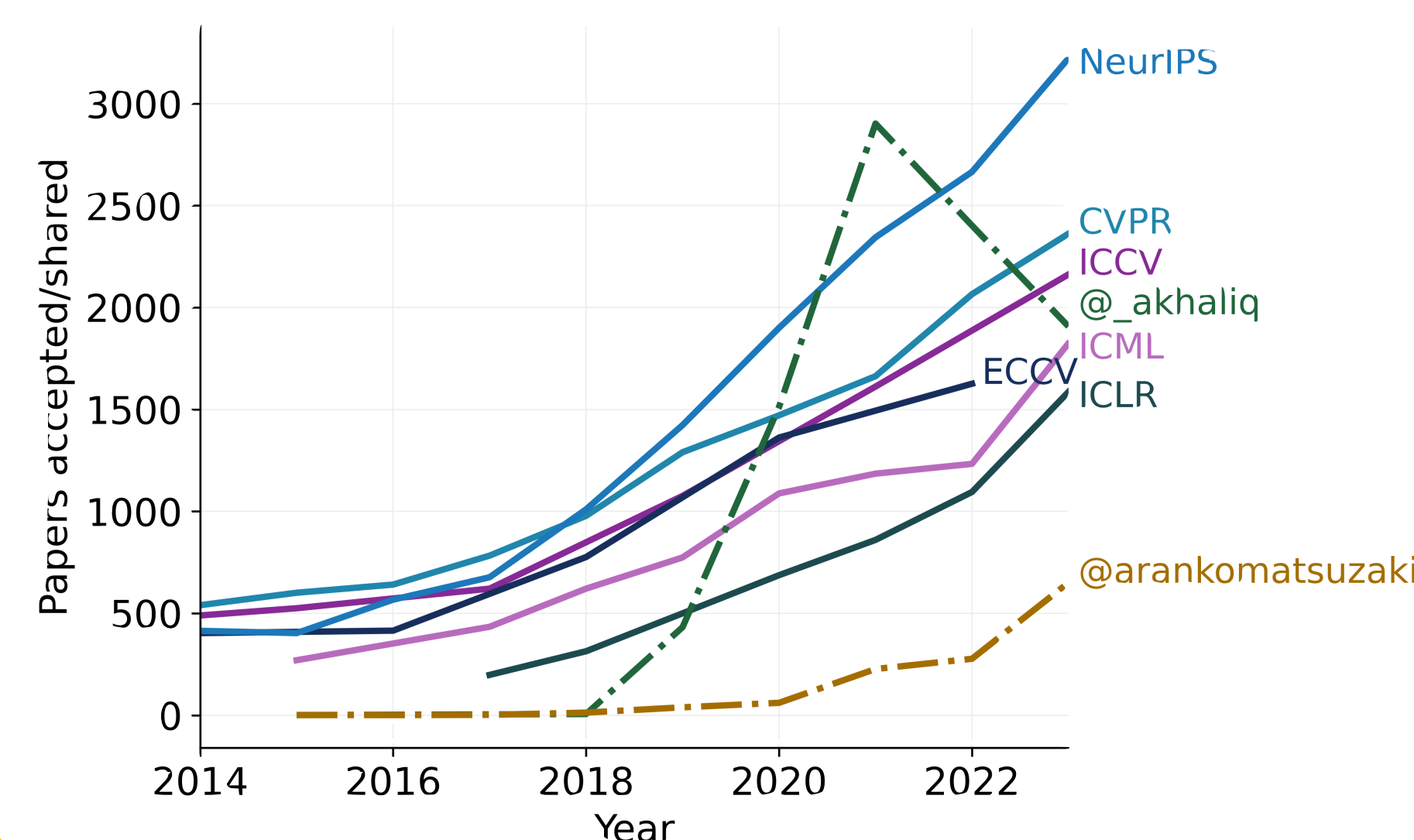We also collect data about **author affiliations, geographic locations, and genders**.

Motivation & Goal

> ### *How do AI/ML researchers find papers?*
> *What concerns does centralized curation bring?*
> *How can the modern conference system adapt?*

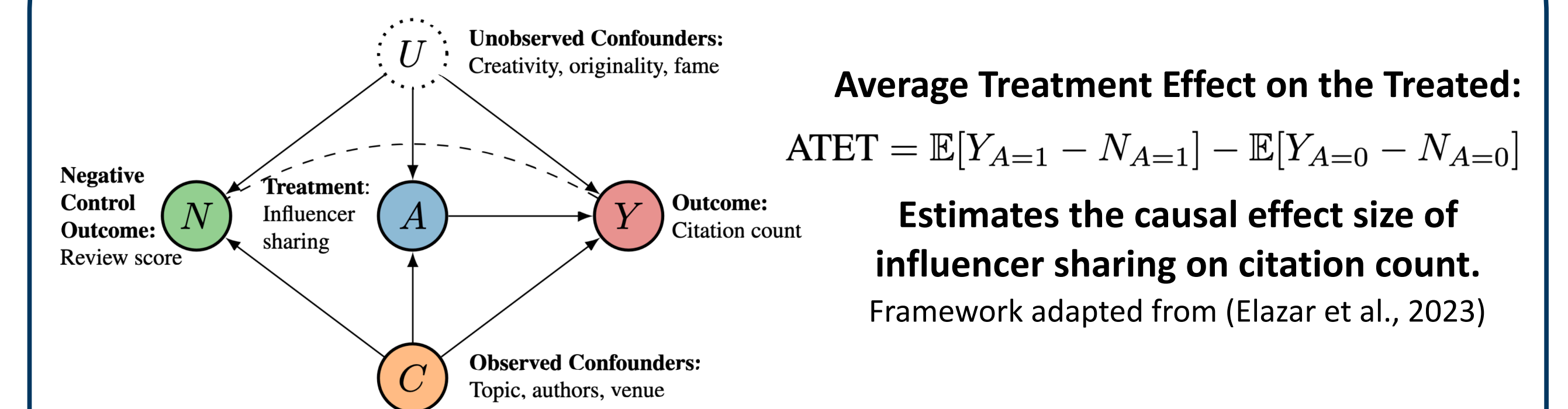### The Explosive Growth of ML Conferences And Influencer Sharing

The rapid growth of AI/ML **makes it harder for researchers to find quality papers**.

They increasingly rely on social media, especially "**Academic Influencers**," to curate relevant papers.

We aim to quantify this trend and **identify ways to improve paper dissemination** and the AI/ML community.

Causal Inference

Unobserved Confounders: Creativity, originality, fame

Negative Control Outcome: Review score

Treatment: Influencer sharing

Outcome: Citation count

Observed Confounders: Topic, authors, venue

Average Treatment Effect on the Treated:

$$ATET = \mathbb{E}[Y_{A=1} - N_{A=1}] - \mathbb{E}[Y_{A=0} - N_{A=0}]$$

Estimates the causal effect size of influencer sharing on citation count.

Framework adapted from (Elazar et al., 2023)

> Influencer sharing has a **significant causal effect** on citation count!

### ATET at 3 Quantile Thresholds w/ 95% CI
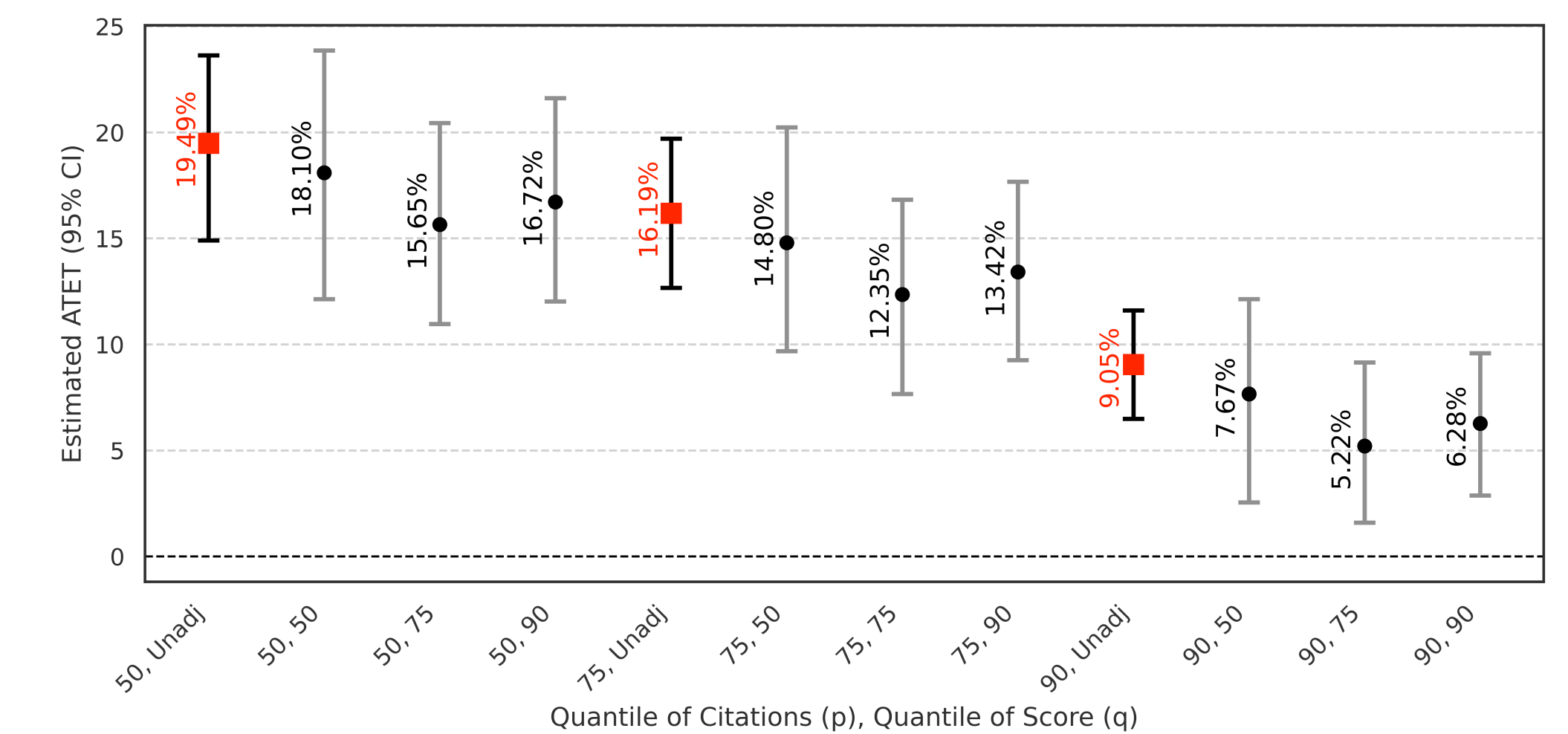
Statistical Analysis

AK
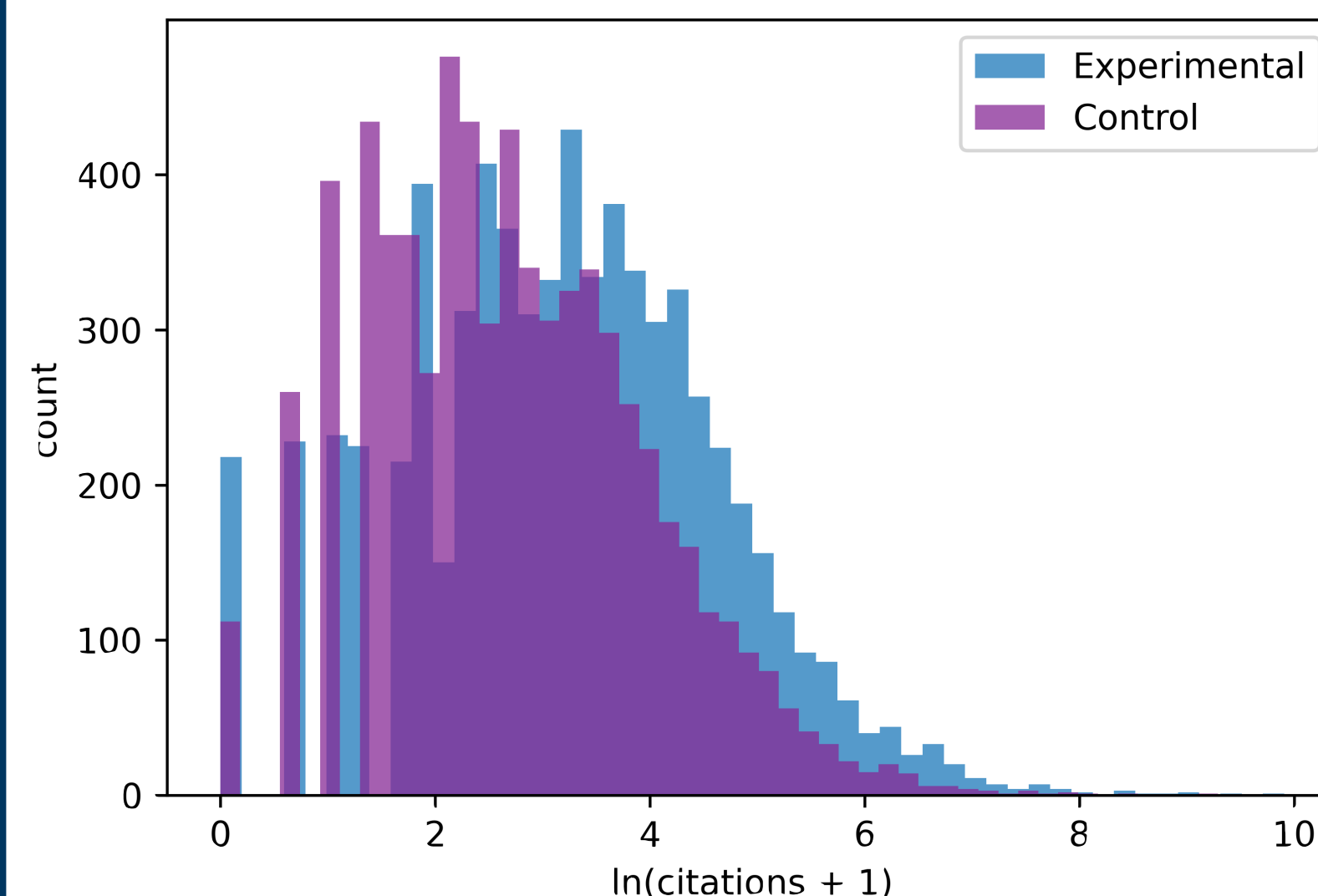
Komatsuzaki

Significance Tests

Test	AK	Komatsuzaki
Epps-Singleton	$p < 0.0001$	$p < 0.0001$
Kolmogorov-Smirnov	$p < 0.0001$	$p < 0.0001$
Mann-Whitney U	$p < 0.0001$	$p < 0.0001$

Treatment distributions are skewed upward from the control.

> Median **citation count is 2-3 times higher** for shared papers.

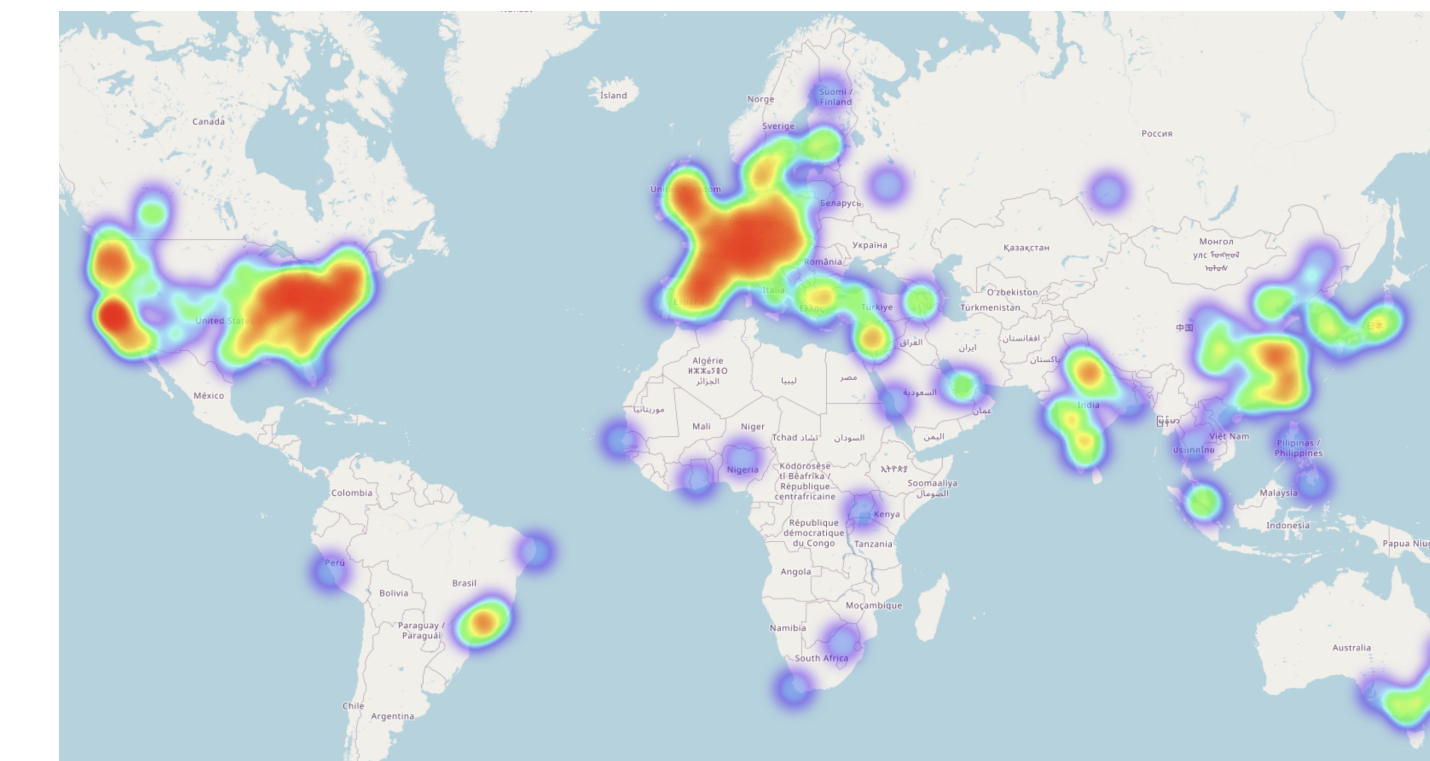There is a **statistically significant increase** in citations for shared papers.
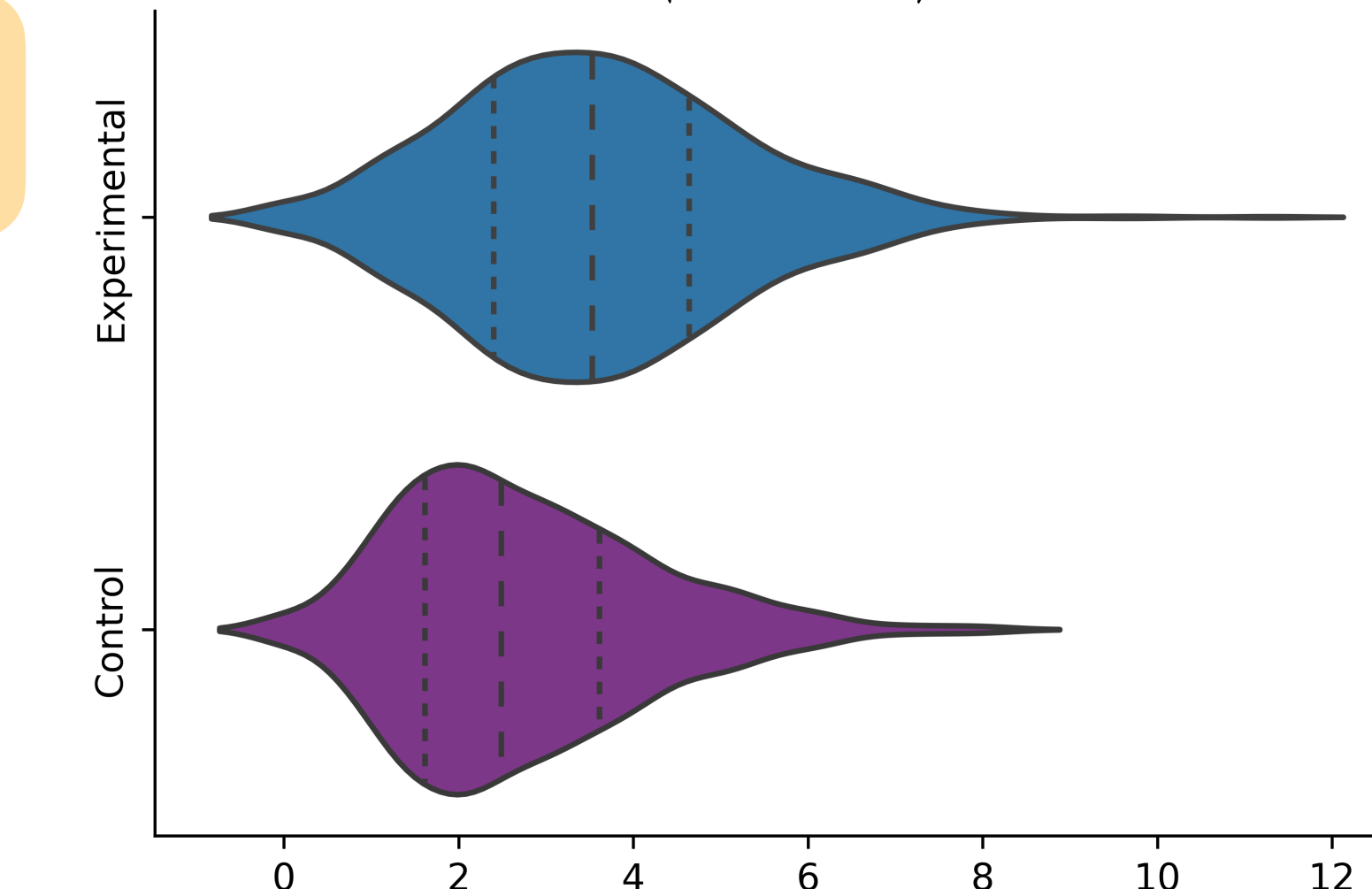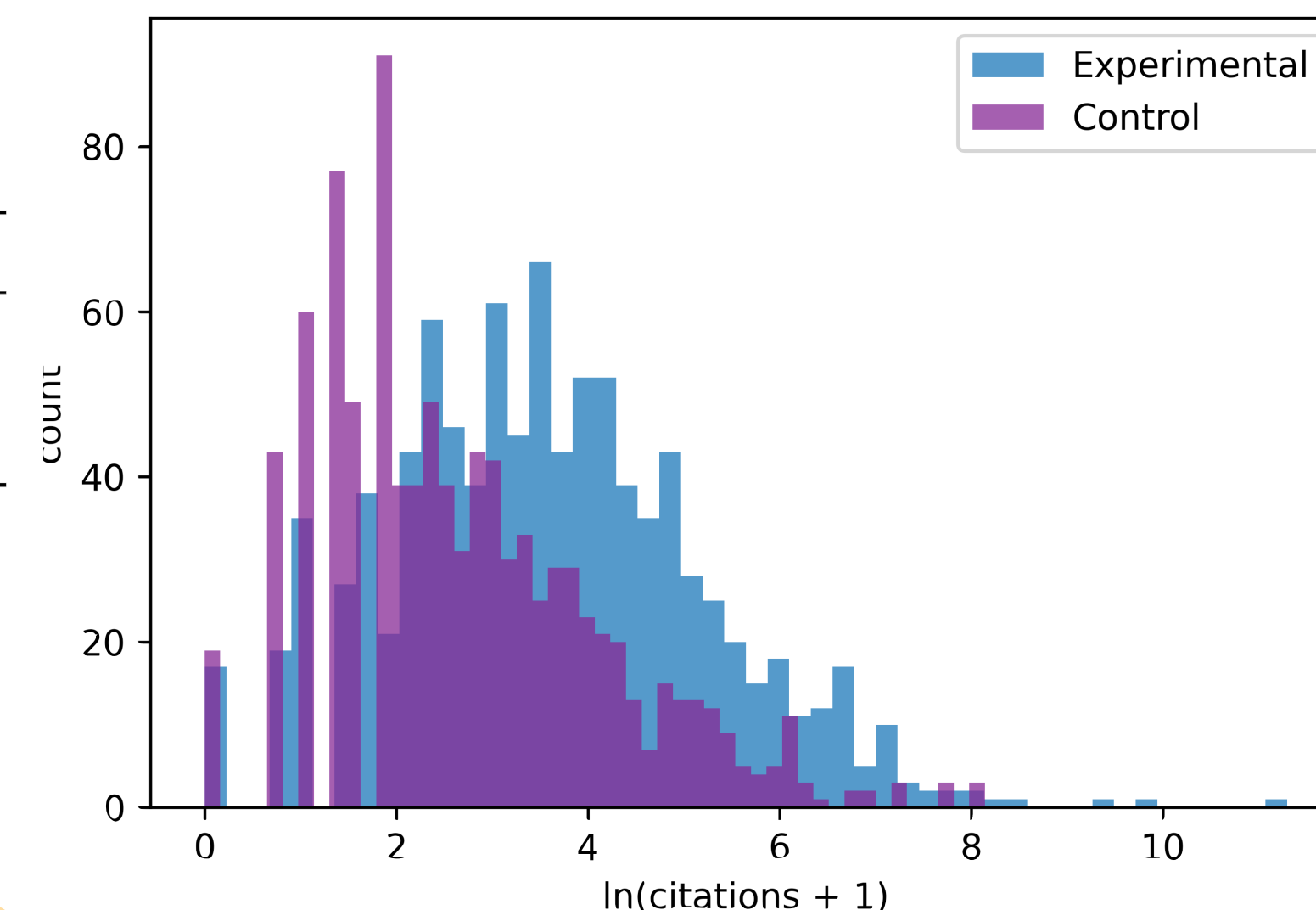
Discussion & Position

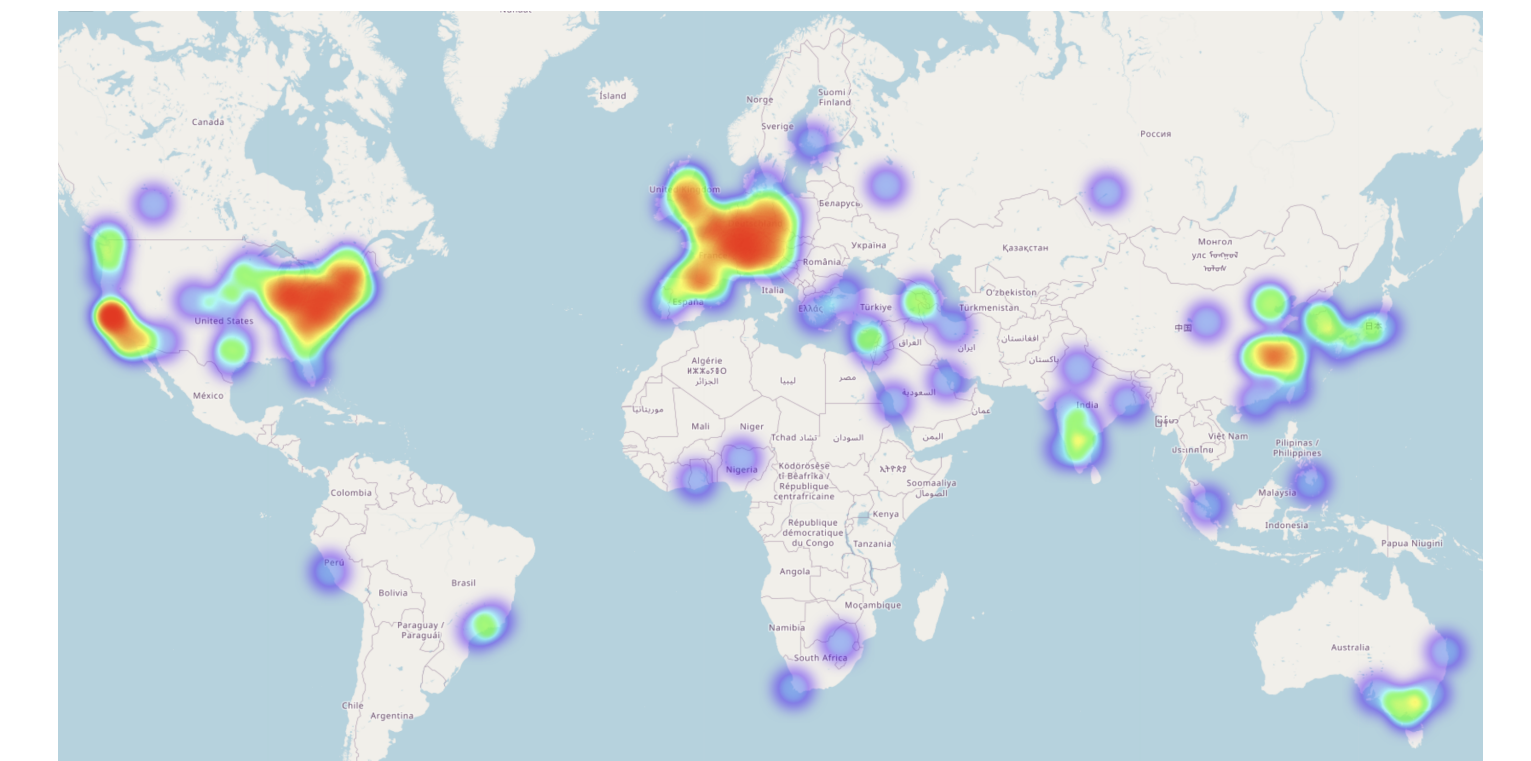### AK Affiliation Heatmap

Komatsuzaki Affiliation Heatmap

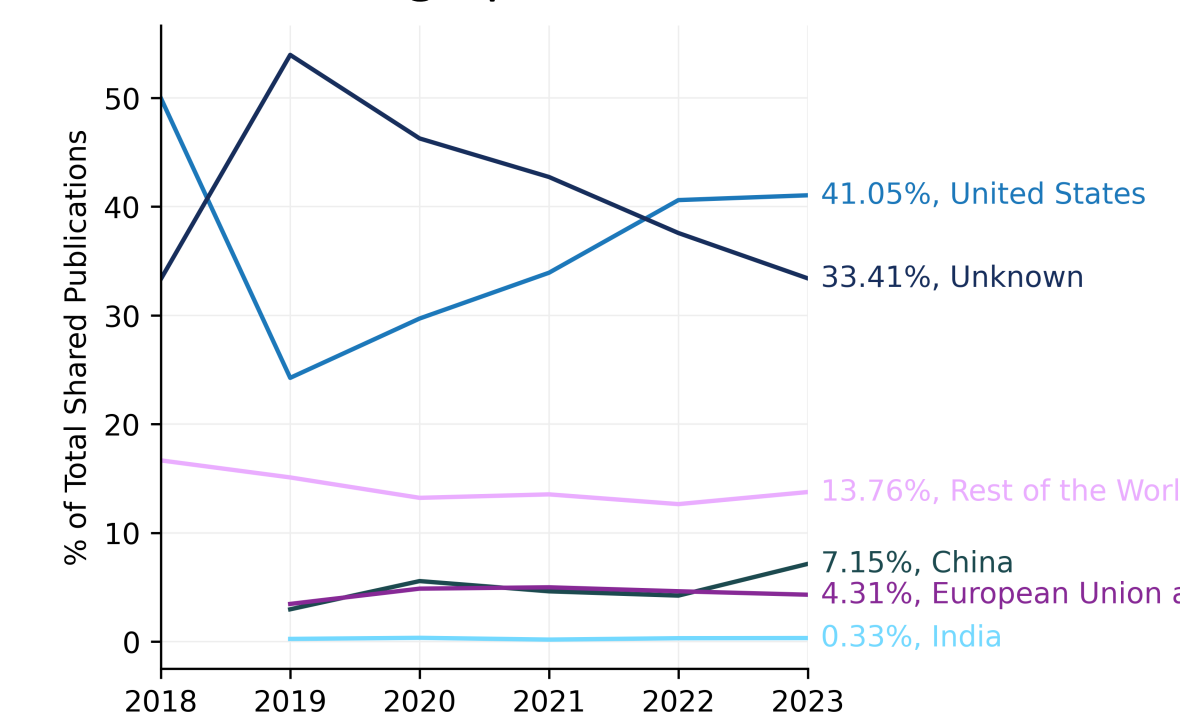While we *do not attribute any intentional bias to our influencers*

We express concerns about the **increasing centralization of the academic process**.

We urge influencers to amplify **diverse ideas and backgrounds** and the AI/ML community to **spot and prevent potential bias**.
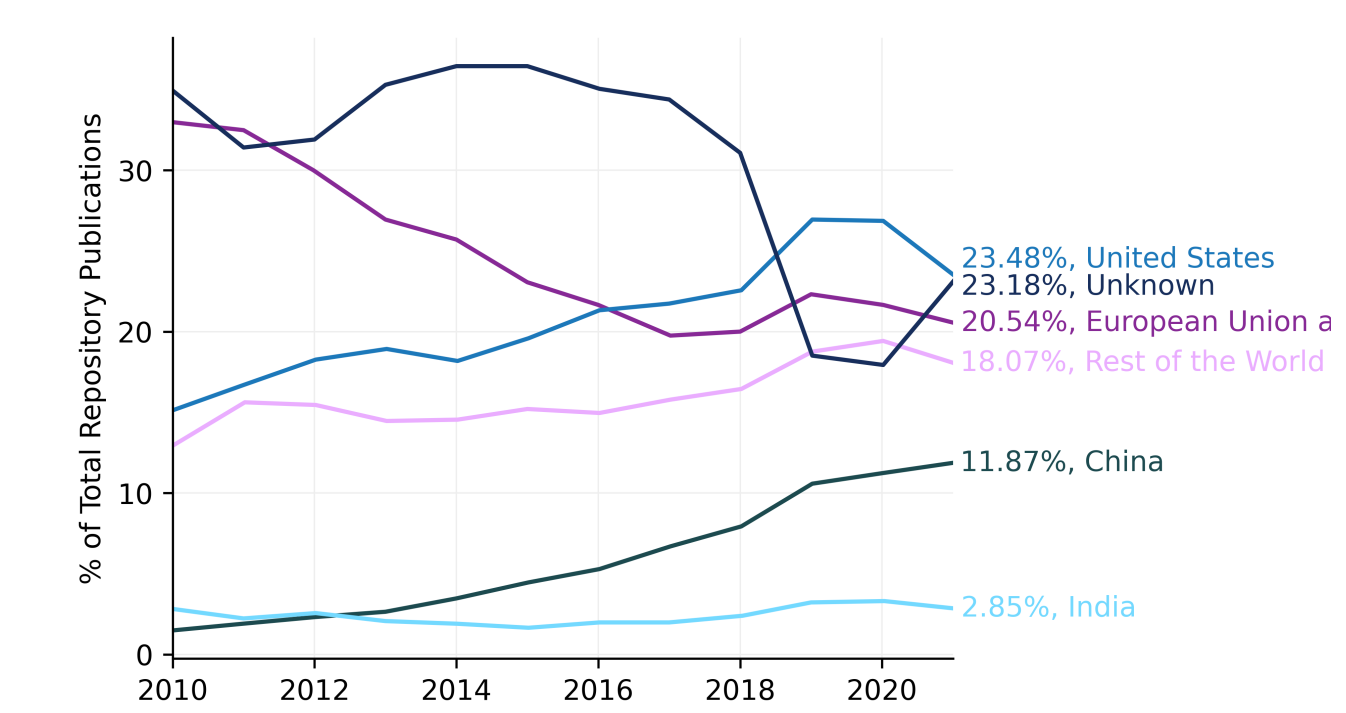
AK Geographic Distribution

Komatsuzaki Geographic Distribution

HAI Report Figure 1.1.19

2023 AI Index Report (Maslej et al., 2023)