

# AI Transparency in practice

What was learnt from third-party audit of  
recommender systems at LinkedIn and Dailymotion.

**AUTHORS**

**JIAHAO CHEN, JACK BANDY, DAVE BUCKLEY, AND RUCHI BHATIA.**

**31 October 2024**

**ISBN:** 978-1-0670444-0-4

**Authors:** Jiahao Chen (Responsible AI LLC), Jack Bandy (Transylvania University), Ruchi Bhatia (HP), and Dave Buckley (independent consultant).

**Contributors:** Jon Adams, Tong Zhou, Adrian Rivera, Alex Murchison, Ryan Rogers, Rahul Tandra, Siyao Sun, Sébastien Le Roux, Thomas Schmitt, Brice de la Brière, Cyrille Brun, Andrew Trask, Irina Bejan, Lacey Strahm, Stephen Gabriel, Zarreen Reza, Ishan Mishra, Osam Kyemenu-Sarash, Laura Ayre and the extremely talented engineering team that shipped and iterated on the platform that facilitated this project to match our research needs.

**With thanks to:** The Christchurch Call would like to extend its thanks to the following people, organisations, and governments for their support, which enabled the delivery of the Christchurch Call Initiative on Algorithmic Outcomes, Phase 1:

The Governments of New Zealand, France, and the United States of America.

Microsoft, OpenMined, Dailymotion, the Center for Security and Emerging Technology at Georgetown University, and Twitter (now known as X).

**Design:** Creature Limited.

Wellington, October 2024

The Christchurch Call Foundation  
P.O. Box 993, Pipitea,  
Wellington, 6011, New Zealand

All rights reserved by The Christchurch Call Foundation. © 2024

This report may not be copied, reproduced, distributed or publicly displayed without reference to The Christchurch Call Foundation or the Report Authors.

# Table of Contents

<b>Summary</b>	4
<b>Introduction to third-party algorithmic auditing</b>	5
<b>Background of CCIAO</b>	6
<b>Challenges working with proprietary data</b>	7
<b>Methods for privacy-preserving third-party audits</b>	8
Remote data science	8
Differential Privacy	9
<b>Data assets</b>	10
LinkedIn data assets	10
Dailymotion data assets	11
<b>Results</b>	13
LinkedIn results	13
Dailymotion results	18
<b>Discussion</b>	20
Benefits and challenges of the research	20
Benefits and challenges of a privacy-preserving audit setup	21
<b>Future work</b>	22
Broadening data access	23
Supporting compliance with platform transparency regulations	23
<b>Conclusion</b>	24
<b>Acknowledgments</b>	25

# Summary

The Christchurch Call Initiative on Algorithmic Outcomes (CCIAO) aims to enable independent study of algorithmic outcomes by addressing data access barriers through new privacy technologies.

This report describes the outcomes of Phase 1 of CCIAO, where OpenMined's PySyft software was used to facilitate external access to impression data related to the production recommender systems at LinkedIn and Dailymotion, and details the research that was then conducted through this platform by four independent researchers (the authors).

We were able to carry out quantitative analysis of the recommender systems at both platforms to answer questions about how these algorithms shape the content recommended to users, whilst protecting the security and privacy of personal or commercially sensitive information through the use of PySyft and the OpenDP differential privacy library.

# Introduction to third-party algorithmic auditing

A rapidly increasing number of people interact daily with AI systems, relying on them for information, decision-making, and daily tasks. How can we ensure these systems are safe and beneficial for people to use?

Third-party algorithmic audits are crucial for AI transparency, as they enable us to identify risks such as unfair bias, intellectual property violations, or the spread of disinformation or other harmful content.

In other domains, findings from third-party audits have led to critical interventions including voluntary moratoria, successful lawsuits, or regulatory changes.<sup>1</sup> However, it can be prohibitively challenging to conduct third-party audits of AI systems due to security, privacy, intellectual property, or trade secret concerns, which prevent external auditors from being able to access and study the key data assets that are consumed or produced by the AI system.<sup>2</sup> Data sensitivity thus inhibits collaboration with external parties, leaving companies to rely solely on internal audits without external feedback or challenge. This lack of oversight can mean that risks go undetected, and harms can persist.

---

1. *Outsider Oversight: Designing a Third Party Audit Ecosystem for AI Governance*, Raji et al, 2022

2. See e.g. *Auditing Work: Exploring the New York City algorithmic bias audit regime*, Groves et al., 2024

# Background of CCIAO

On March 15, 2019 a terrorist attacked two mosques in Christchurch, New Zealand, killing 51 people and injuring 50. The horrific event was live-streamed on Facebook by the attacker for 17 minutes. Although fewer than 200 people watched the original, live broadcast, copies of the video spread rapidly to multiple platforms including Twitter, Youtube, and Reddit, ultimately [reaching millions of viewers](#). As part of the response to the attack, In response, New Zealand's then Prime Minister Jacinda Ardern and French President Emmanuel Macron brought together Heads of State and Government and leaders from the technology sector to adopt the Christchurch Call, with the goal to "eliminate terrorist and violent extremist content (TVEC) online". The [Royal Commission Inquiry](#) into the attack later concluded that the terrorist was radicalised, in part, by content found online through social media platforms, further emphasising the importance of the Call's mission.

The Call Community comprises 56 governments, 19 online service providers, 12 partner organisations, and a Christchurch Call Advisory Network of more than 50 civil society organisations and individuals.

To achieve this ambitious goal, we first need a deeper understanding of the role that algorithms and AI systems play in the spread of TVEC online. However, as previously mentioned, numerous access challenges have acted as restrictions on carrying out the required research. To overcome this, in September 2022, In September 2022, in conjunction with the UN General Assembly and the Christchurch Call Leadership Summit, Jacinda Ardern announced the [Christchurch Initiative on Algorithmic Outcomes \(CCIAO\)](#) The initiative specifically aims to accelerate technology development to enable independent study of algorithmic outcomes by addressing data access barriers, facilitating research that is reproducible, scalable and affordable.

We reported on initial findings in Paris in November 2022, with a goal to develop and test methods for third-party researchers to audit a proprietary algorithm. The remainder of this report describes these methods and how they were applied to audit recommender algorithms at LinkedIn and Dailymotion.

# Challenges working with proprietary data

Traditional audits of a proprietary system require that the auditor a) obtains a copy of the data/model/software, b) goes on-site to have direct access, c) uses an API the company created. Whilst these approaches can be successful, they are not without limitations:

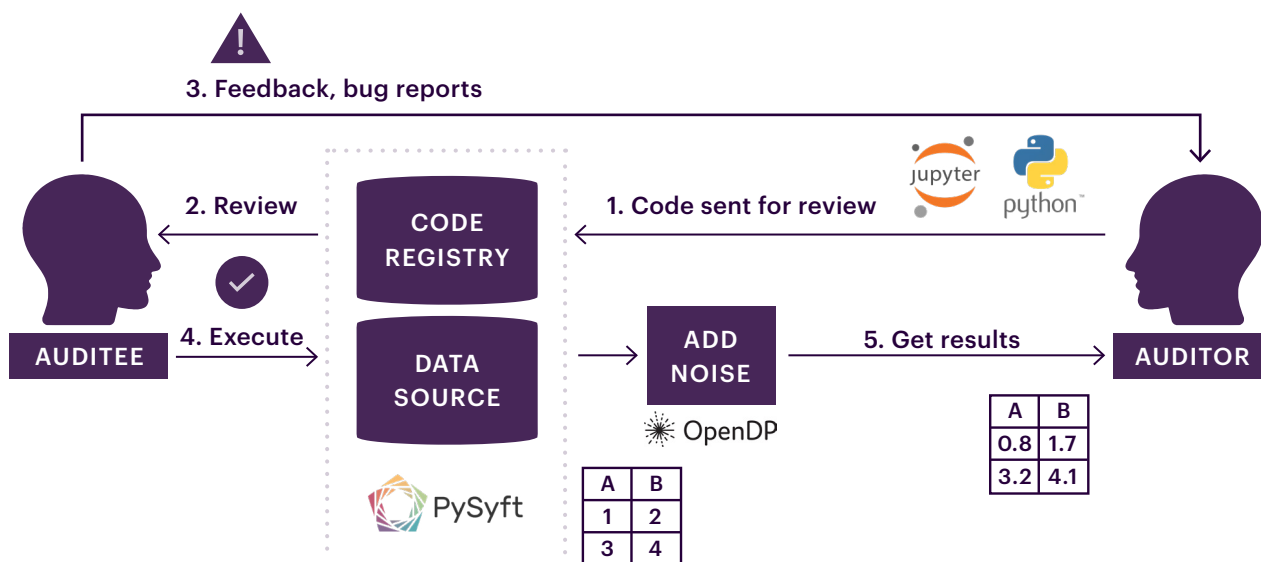
- **Option a)** potentially requires access to large compute resources and engineering time to re-implement the system.
- **Option a)** introduces the possibility of drift between the production system and the system being audited.
- **Option a)** exposes the system to misuse, as the data owner does not have control over what the external researcher then uses the system for, or who they share it with
- **Option b)** potentially limits the level of access the auditor has, as they are subject to the use policies and resource constraints of the platform.
- **Option b)** places significant limitations on who can audit the system, as the costs associated with bringing auditors on-site are high.
- **Option c)** grants companies the power to limit audit types, in that companies could define what types of audits their APIs support and build custom APIs only allowing what they deem as permissible audits. In the event an audit requires information outside of the custom-built API the company already created, the company could deny the request on the basis of resource costs.

Less traditional methods of data release that respond to some of these limitations include the release of synthetic data. In particular, differential privacy was deployed in releasing data from the 2020 US Census to “formally “ protect the confidentiality of users, at the trade-off of data quality and utility. However, such solutions are known to have a disparate impact on minority groups and the synthetic data generation needs to be tailored to a specific type of data analysis to ensure robustness.

# Methods for privacy-preserving third-party audits

CCIAO aims to overcome the limitations of these approaches by using new oversight tools that enable *privacy-preserving* audits. The purpose of Phase 1 was to build audit infrastructure that leverages these tools, which we – serving as independent external researchers – could then test to assess their feasibility.

In principle, privacy-preserving audits allow third parties to study proprietary systems without direct access, thus mitigating risks related to privacy, security, and intellectual property. We have piloted PySyft, an open-source library that implements a flexible approach based on **remote data science**, in which data owners decide to allow third parties to remotely query relevant datasets. This approach allowed us to integrate other third-party privacy-preserving tools, such as **differential privacy**, which helps ensure that the outputs of queries preserve user anonymity.



## Remote data science

Remote data science helps enable third-party auditing because data owners maintain control over their datasets during the audit process. Raw data never leaves the data owner's infrastructure, and the data owner retains the authority to approve or deny auditors' queries, as well as validate the researcher's findings. This is the paradigm that PySyft implements.

To address key questions, we were able to prototype the required analysis on mock datasets – data with the same schema as the real data, but fake data entries – and test our auditing code. This mock data supports debugging and other iterative steps in the data science workflow, which the auditor can continue without waiting for approval. Upon submission and approval on the real data source, we were able to fetch the answers to our analysis and request follow up analysis to conclude the audit.



## Differential Privacy

Differential privacy represents another layer in the privacy-preserving auditing stack. This layer helps ensure that any results retrieved from a dataset (by an auditor) protect the privacy of people in the dataset. In this pilot, we use a differential privacy mechanism that adds a controlled amount of noise to the original dataset while ensuring that results are “close enough” to the original dataset. Use of differential privacy was not a requirement from all data owners participating in this pilot, as data release policies differed across organisations. When it was required, we used the open-source [OpenDP](#) framework within PySyft.

One factor to consider when using differential privacy is the tradeoff between privacy guarantees and precision of results. In practice, a privacy budget is used to limit how much noise is added, balancing privacy and accuracy, which represents the cumulative amount of noise used across multiple queries. This allows auditors to spend more of their privacy budget when they need more precision.

# Data assets

Both data providers made available impression data about their recommender systems – this is useful for auditing since it captures the frequency and target audience for specific social media content, helping us to understand algorithmic outcomes, the reach of content to various user groups, and potential biases in the recommendation process.

Today, only a few platforms release impression data for purposes of auditing and transparency. For example, Facebook releases [four transparency reports](#) per year which include helpful information about widely-viewed content. CCIAO aims to build on this kind of transparency to allow more frequent and larger-scale auditing.

## LinkedIn data assets

Variable name	Data type	Description
<b>posterId</b>	integers	Internal id
<b>viewer_industryCategory</b>	categorical labels (18)	Industry in which the viewer is working
<b>isShareJobOpportunity_list</b>	Boolean	Whether or not the post shares a job opportunity to the viewer
<b>viewer_hasAtLeastCollegeDegree</b>	Boolean	Whether the viewer has a profile that lists a college degree or higher level of education.
<b>isAIRanked</b>	Boolean	The algorithm that ranked the post (true and false label two different algorithms)

Table 1: Schema for the LinkedIn dataset.

The LinkedIn dataset consisted of ~70 million rows of data related to LinkedIn public post activity. Each row in the dataset represents the top-ranked post in a user’s feed for a particular session. Information is provided on which algorithm was used to rank the post (**isAIRanked**), whether the post is sharing a job opportunity (**isShareJobOpportunity\_list**), and information about the user’s industry (**viewer\_industryCategory**) and education (**viewer\_hasAtLeastCollegeDegree**). The schema is shown in Table 1.

	<b>count</b>	<b>viewer_industryCategory</b>	<b>isShareJobOpportunity</b>	<b>viewer_hasAtLeastCollegeDegree</b>	<b>is AIRanked</b>
0	51273	Recreation, Travel, and Entertainment	True	None	True
1	63519	Manufacturing	False	False	False
2	361894	Legal	None	True	True
3	2366610	Manufacturing	False	True	True
4	111081	Construction	True	True	True
...	...	...	...	...	...
319	1847	Service Industry	True	None	False
320	34695	Organizations and Nonprofit	True	True	True
321	22194	un	None	True	False
322	543	un	True	False	True
323	43207	Government	True	True	True

324 rows x 5 columns

Table 2: differentially-private 4-way contingency table for the LinkedIn data, derived by aggregating posts over the viewer\_industryCategory, isShareJobOpportunity, viewer\_hasAtLeastCollegeDegree, and isAIRanked variables.

Differential privacy was crucial for enabling privacy-preserving data release in this context. With a privacy budget capped at a given value by LinkedIn's data managers, we explored building a 4-way contingency table (see Table 2) to allow us to conduct our analysis without exceeding the budget, benefiting from the immunity to post-processing of differential privacy. This allowed us to overcome the challenge of manually managing the privacy budget spending across various analyses.

We framed our investigation of the LinkedIn data around two research questions:

- **RQ1:** Are there detectable patterns of post viewing activity based on user demographics?
- **RQ2:** Does the choice of algorithm affect the likelihood of a user being served a job ad?

### **Dailymotion data assets**

For Dailymotion, data was provided for ~10 million records of videos. Each row in the dataset represented impression data about a video that is part of the platform and metadata about which of three algorithms was used to rank it, the number of times the video was recommended by the algorithm, and a so-called "suggestiveness score" determined by an internal scoring model summing up the degree of violent or sexual content in that video.

In studying the Dailymotion data, we sought to address the following two research questions:

- **RQ3:** Do the different algorithms promote suggestive videos at different rates?
- **RQ4:** How do the different algorithms impact the equality of post promotion?

# Results

In Phase 1, all four of us – acting as independent third-party auditors – were able to successfully perform queries and derive results using the privacy-preserving tooling that had been built.

## LinkedIn results

### RQ1: Are there patterns of post viewing activity based on user demographics?

Firstly, we investigated how viewing activity varied across industries. A baseline for comparison was created using US employment data taken from the Bureau of Labor Statistics website (CPS). The data for non-agricultural, for-profit employers is taken from the May 2023 of Current Employment Statistics, [Table B-1a](#) (seasonally adjusted figures), using the NAICS codes in the following table. Agricultural data is taken from the 2021 figure in [Table 2.1](#).

BLS does not report nonprofits and for-profits separately. Instead, the employment data on nonprofits comes from the most recent special report on [nonprofit employment](#) (2017). The employment rates in all sectors for CES are scaled down proportionally to estimate the proportion held by for-profit employers.

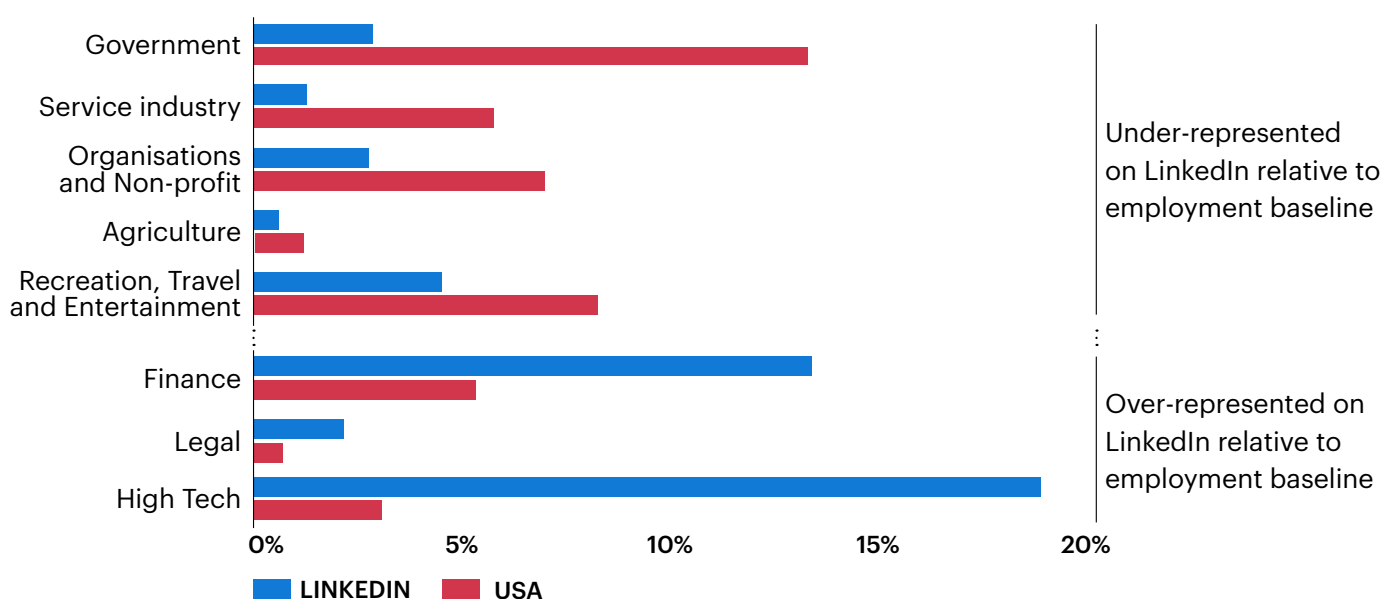


Figure 1: Percentage of LinkedIn posts per industry category (blue), compared to a baseline derived from US employment data from the Bureau of Labor Statistics (red).

Figure 1 shows the percentage of LinkedIn posts per industry category, compared to this baseline. The results show that High Tech, Legal, and Finance are the most overrepresented industries on LinkedIn – this is perhaps consistent with what one might expect, with workers in knowledge work industries being heavy users of LinkedIn. At the other end of the spectrum, the most underrepresented categories include sectors that typically involve more manual labor and/or less computer-based work (e.g. Service Industry, Agriculture). Government is ranked as the most underrepresented sector on LinkedIn, which we speculate could be due to strict social media policies placed on government employees.

These results must be interpreted with caution due to the fact that the baseline was derived using employment data for the US, whereas the LinkedIn data is not geographically restricted. This exemplifies a more general challenge of deriving suitable baselines in social media research, which is discussed in more detail below.

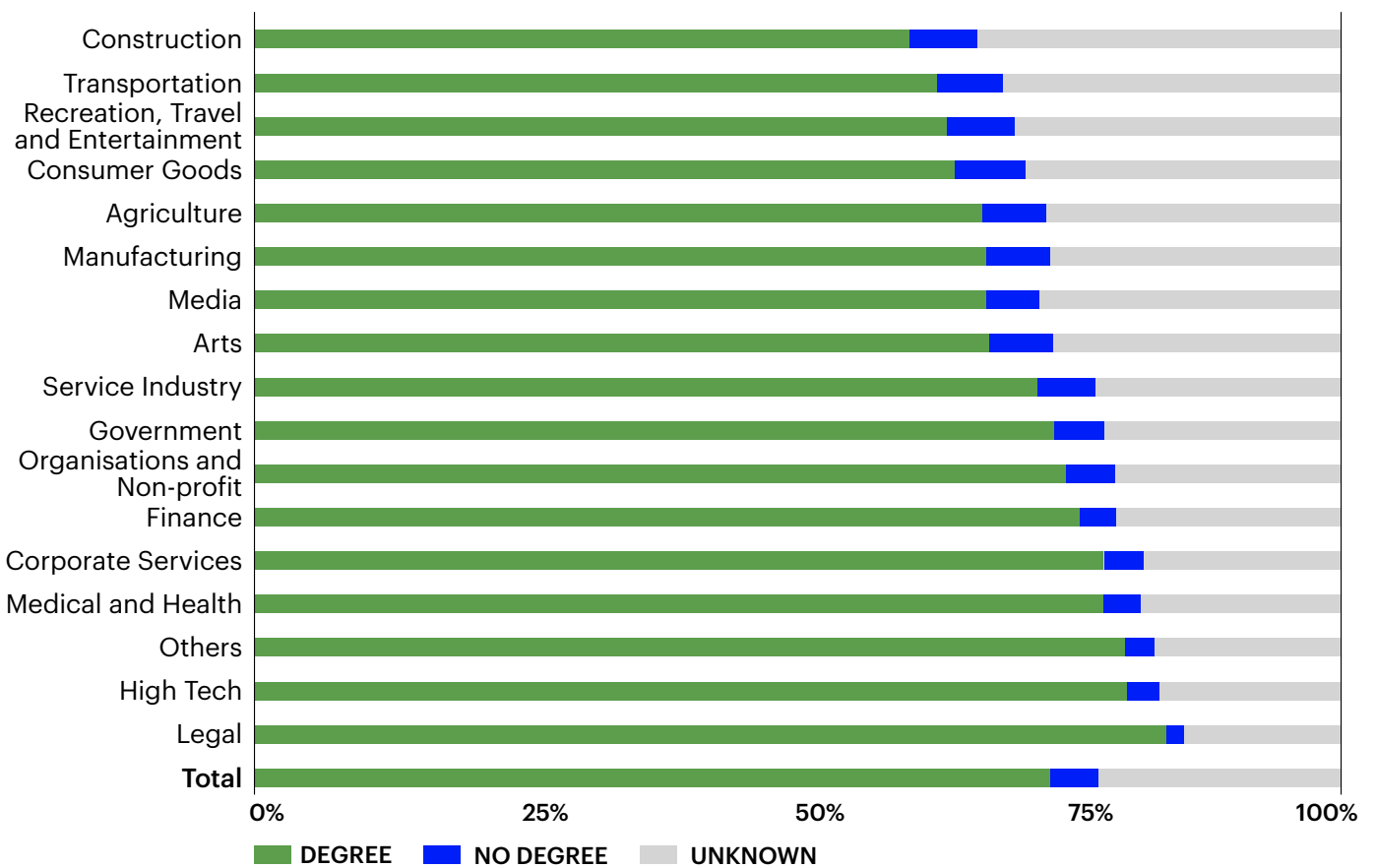


Figure 2: Percentage of users who hold a degree (according to their LinkedIn profile) across each industry category.

We also find (see Figure 2) that users in different industries have different levels of educational attainment. Again, we see patterns that we might anticipate with users in the Legal, High Tech, Medical and Health professions more likely to have a degree. However, we must again be cautious in our interpretation of the results as in this context “has a degree” really means “has a degree listed on their LinkedIn profile”.

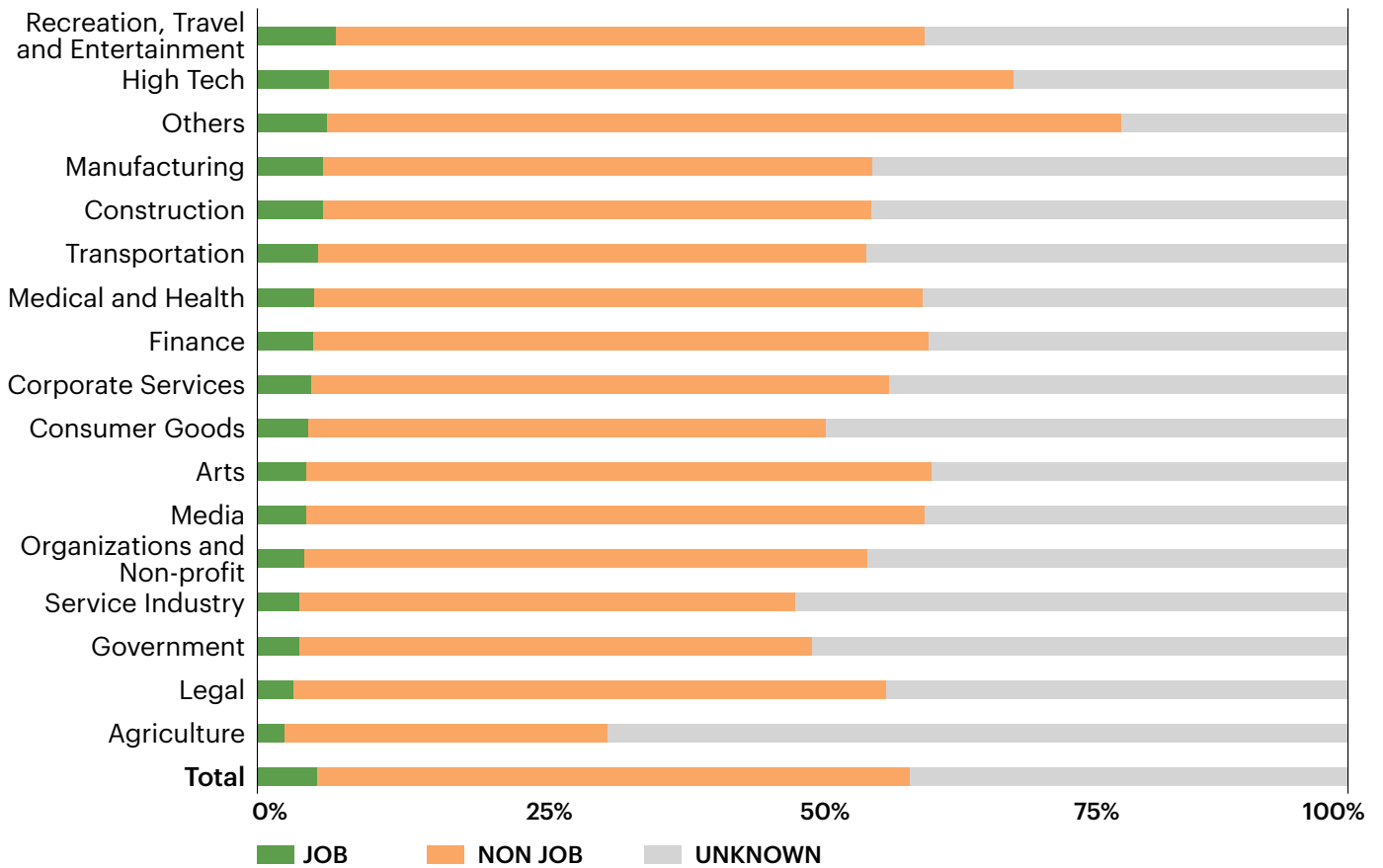


Figure 3: Views of posted jobs across each industry category.

Finally, we find (see Figure 3) that views of posted jobs varied across industries.

**RQ2: Does the choice of algorithm affect the likelihood of being served a job ad?**

Here, we investigate whether the two different ranking algorithms affect the probability of a user being served a job ad.

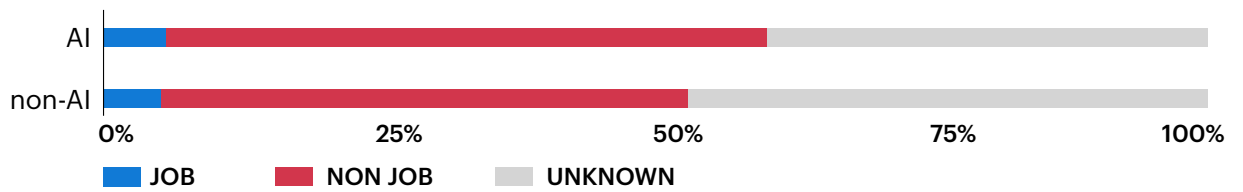


Figure 4: Percentage of highest-ranked posts that served a job ad, for the two different ranking algorithms (AI and non-AI)Figure 5: Percentage of highest-ranked posts that served a job ad, for the two different ranking algorithms (AI and non-AI), across each industry category.



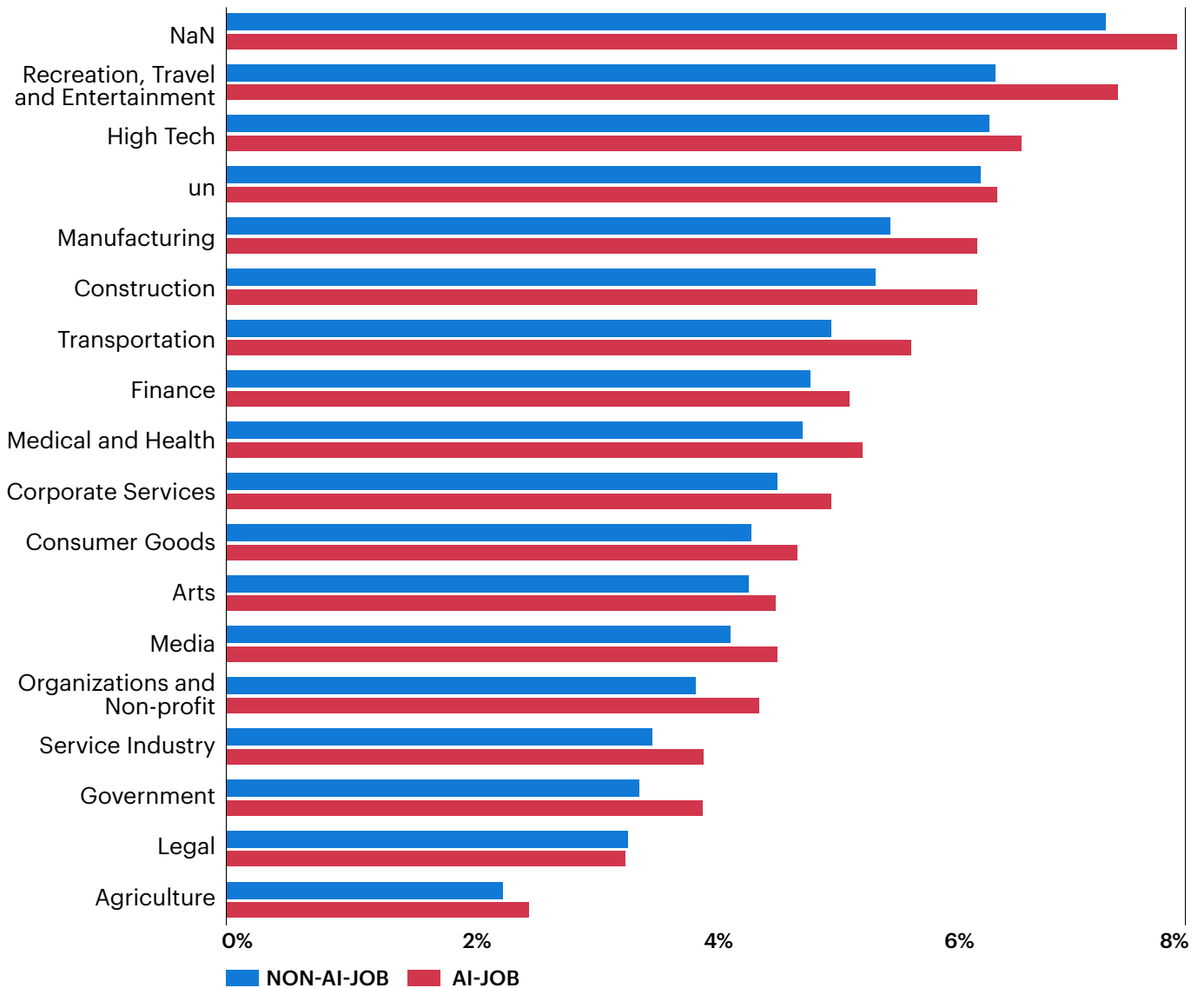


Figure 5: Percentage of highest-ranked posts that served a job ad, for the two different ranking algorithms (AI and non-AI), across each industry category.

Figure 4 shows that the AI algorithm was marginally more likely than the non-AI algorithm to rank a job-related post highest. This was found to be the case (Figure 5) across all industries except Legal.

## Dailymotion results

RQ3: Do the different algorithms promote suggestive videos at different rates?

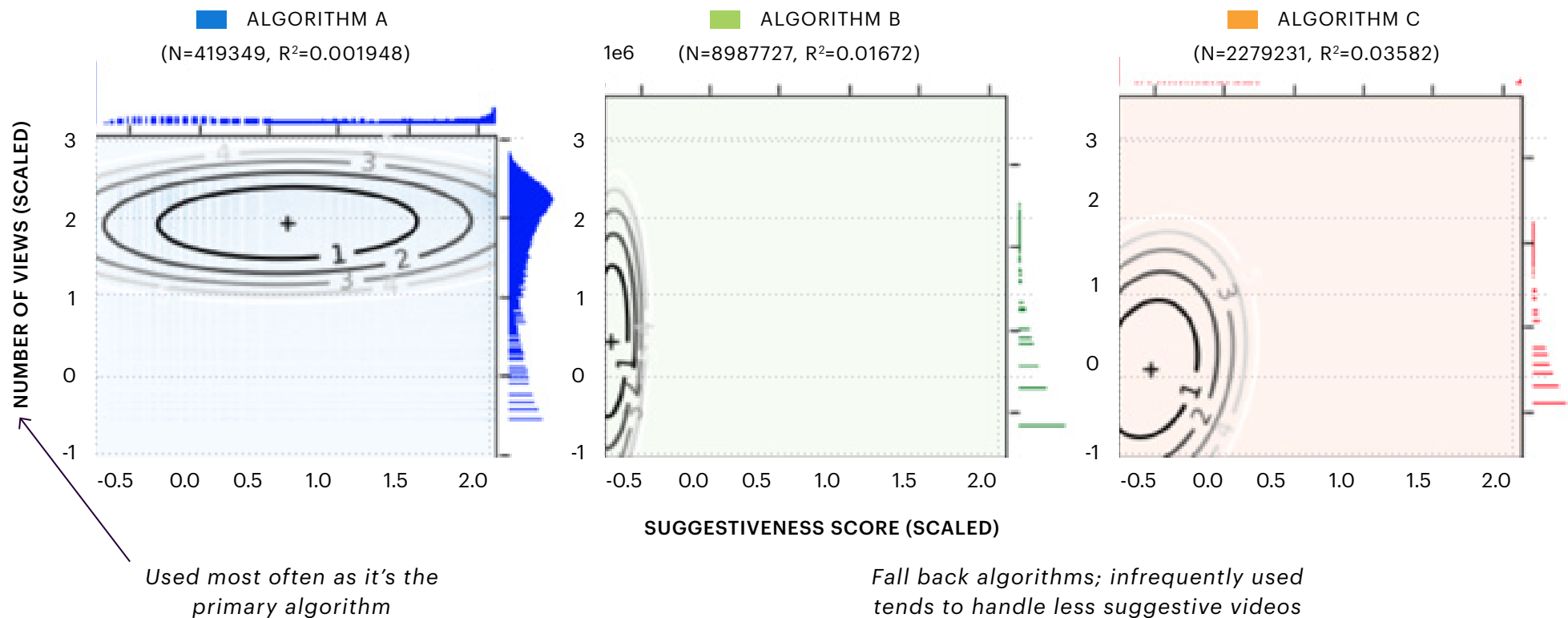


Figure 6: Plots showing the (scaled) suggestiveness score and number of views for a Dailymotion video for the three different recommender algorithms. Algorithm A promotes videos across the whole range of suggestiveness scores, whereas Algorithms B and C only appear to handle videos with low suggestiveness scores.

A surprising output of the analysis was that the algorithms had drastically different distributions in regards to the suggestiveness score of the content that was promoted (see Figure 6). The primary algorithm used tends to promote videos across the whole range of suggestiveness scores, suggesting that it does not account for suggestiveness in particular. The other two algorithms, which are infrequently used as fall-back algorithms, are promoting videos with significantly lower suggestiveness scores. While domain knowledge is required to understand these effects, a challenge in constructing a robust analysis came from the realism of the mock data, which implied a normal distribution, thus a few iterations were required to understand the specifics of the data and required close collaboration with the data owner.

#### RQ4: How do the different algorithms impact the equality of video promotion?

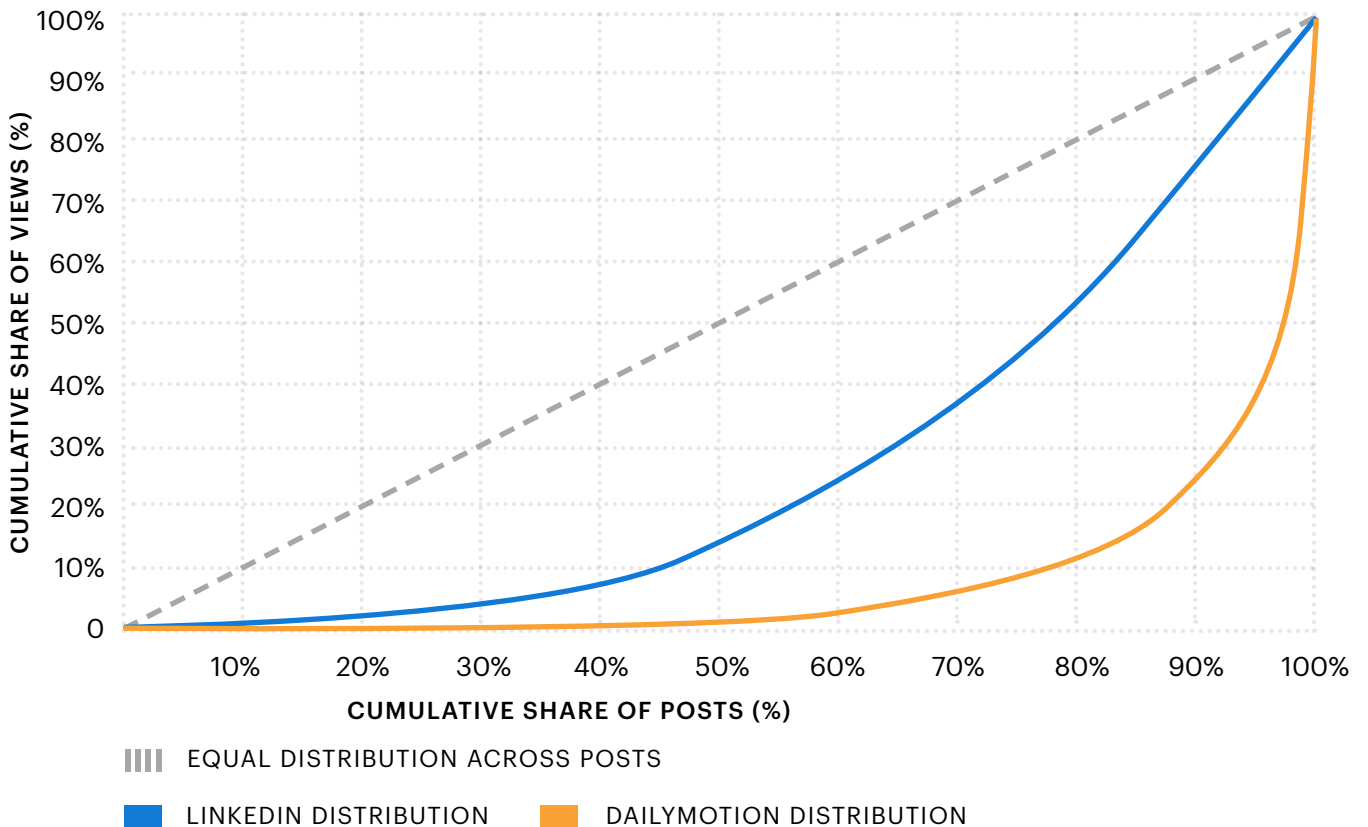


Figure 7: Lorenz curves for the distribution of post views from a LinkedIn algorithm and a DailyMotion algorithm. Notably, the DailyMotion algorithm distributes views with a high degree of inequality, with the top 10% of videos accounting for 75% of all recommendations.

Another line of research developed an algorithm for calculating the Gini coefficient under differential privacy guarantees. The Gini coefficient is calculated from Lorenz curves, as shown in Figure 7, and can be used as a metric to measure inequality within algorithmic systems. Statistical inequality in this context refers to the uneven distribution of recommendations made by an algorithm for different posts. For example, Figure 7 shows that views from the main Dailymotion algorithm exhibit more inequality compared to LinkedIn’s algorithm. Under Dailymotion’s algorithm, the top 10% of video posts accounted for 75% of all recommendations. LinkedIn’s algorithm distributed views more equitably, with the top 10% of posts accounting for closer to 25% of all recommendations.

# Discussion

## Benefits and challenges of the research

These results show that auditors were able to successfully run analyses against the production recommender systems and derive substantive results. We gained insights into viewing habits and recommendation behaviour across different demographic groups, and studied the impact of different algorithms on content recommendation.

To our knowledge, this is one of the first cases in which external researchers remotely analysed impression data from social media at this level of detail and with active collaboration from the auditee. Prior initiatives from within companies, such as CrowdTangle, Social Science One, and the Twitter Academic API, have seen mixed success. For example, when using Facebook data through Social Science One, “Researchers should not be computing the ratio of variables with noise”<sup>3</sup> (among other challenges and measurement errors). The “structured transparency” approach we describe offers more flexibility to data scientists, allowing them to combine multiple methods to preserve privacy rather than relying on companies to choose their own methods.

However, the research was not without its challenges and limitations. A key challenge was establishing a suitable baseline for comparison. This is important for discerning the direction of an algorithm’s influence – is it making things better, making things worse, or maintaining the status quo? In the context of LinkedIn job posts, compelling baselines were available through the U.S. Bureau of Labor Statistics. In other contexts such as audience sizes on social media, relevant baselines are not as accessible from legacy media. We suggest further work in the following section to combat this challenge.

It should also be noted that data access alone is not necessarily sufficient for an auditor to be able to make conclusions about the behaviour of the system – often domain-specific knowledge is required. We saw this in the RQ3 where it transpired that different algorithms were being applied in different contexts, a fact that we were only able to learn through conversations with Dailymotion. This is a crucial lesson: effective external audit of a system requires effective interaction and close collaboration between the auditor and system owner.

---

3. From Facebook’s Frequently Asked Questions for the SS1 researcher platform

## Benefits and challenges of a privacy-preserving audit setup

Remote data science provided benefits to both auditors and data owners. For us the auditors, the privacy-preserving features of the platform meant that we could carry out our work at any time from any location with an internet connection (3 of us were based in the US, 1 in the UK). Additionally, this setup reduced our liability relative to more traditional approaches since the possibility of misuse was minimised by the fact that we were never able to see the real data, and any code requests had to be reviewed and approved by the data owner before being executed.

The combination of remote data science (an input privacy technique preventing access to the underlying data) with differential privacy (an output privacy technique preventing inference of sensitive information from the output of the analyses) provided the data owners with strong, provable guarantees over the end-to-end privacy of their data assets. Moreover, integrating a manual approval process provided the assurances that the researchers were not attempting to run code that could be considered as misuse. Whilst this ‘belts and braces’ approach was effective, it should be recognised that it will not always be appropriate. For example, there are likely scenarios where it may be in a regulator’s interest to keep auditing code private from a company, to prevent them from “gaming” the evaluation. The use of additional privacy technologies such as secure enclaves, coupled with appropriate regulatory requirements, could facilitate such a use case.

Finally, the use of PySyft and OpenDP introduced new technical challenges that we had to work through. Firstly, the requirement to implement differential privacy in our auditing code and suitably allocate privacy budget proved initially challenging, though (for the LinkedIn case) it turned out that we were able to simply allocate all our budget to the generation of the contingency table, and carry out further exploratory analysis on this table client-side.

Secondly, we met unforeseen issues when transitioning from using the local mock data to the remote production data. This was primarily due to the mock data being smaller in size, and having a ‘cleaner’ distribution that was Gaussian with few outliers (cf. the production data which was non-Gaussian with long tails and many outliers). This required an iterative process to learn about the data from subsequent requests to construct the analysis, as the mock data properties proved to be sometimes misleading. It is therefore important that auditors develop code in a way that is able to anticipate these failure modes.

# Future work

## **Broadening data access**

Whilst in Phase 1 of CCIAO we were able to prove that effective analysis of proprietary algorithms can be carried out using PySyft, the scope of our research was limited by the data made available. Future work should expand access to a broader set of data, increasing both the volume and diversity of data available. Specifically, providing access to data that is directly related to TVEC is vital for fulfilling the goals of the Christchurch Call.

Research into TVEC and other online harms will also benefit from access to data from multiple platforms. This will help auditors gain a more comprehensive understanding of how different algorithms affect content recommendation, and how content spreads between platforms. Furthermore, enabling access to platform data alone is likely insufficient for understanding the real-world impacts of harmful online content. Future work should therefore consider providing access to additional data sources, for example social care data to better understand the link between social media usage habits and mental health outcomes. Access to such auxiliary datasets could also provide data that helps auditors to create more realistic benchmarks.

Future work will also perform more fine-grained analysis regarding different types of content. For example, content about politics may exhibit different exposure and engagement patterns compared to, for example, content about pets.

## Supporting compliance with platform transparency regulations

In recent years, we have seen several jurisdictions introduce legislation aiming to mitigate online harms and place greater accountability on platforms. This includes the introduction of provisions explicitly requiring platforms to make internal data available to regulators and/or external researchers. For example, [Article 40](#) of the European Union's Digital Services Act enables the regulator to demand designated platforms make data accessible for research by vetted researchers. In the UK, [S. 100\(3\)](#) of the Online Safety Act provides the regulator with legal powers to run and observe empirical tests against the algorithmic systems of designated service providers. In the US, [S.1876 – the Platform Accountability and Transparency Act](#) was introduced in the Senate and would support research about the impact of digital communication platforms on society by providing privacy-protected, secure pathways for independent research on data held by large internet companies.

We welcome the introduction of transparency legislation, but recognise the challenges of making their accompanying regulation effective in practice. We believe that Phase 1 of CCIAO demonstrates a promising way forward. Previous transparency paradigms and their corresponding regulations relied on the premise that researchers required a copy of the raw data.

Our work with the CCIAO demonstrates a progression in this paradigm; transparency can exist without copying, sharing, or transmitting the raw data, as organisations can reclassify such research activities as releasing information about the raw data rather than permitting access directly to it. Therefore, valuable research that may previously have been prevented by regulations that limit the sharing of sensitive data can instead safely proceed. Future work is encouraged between the Christchurch Call, platforms, and regulators to explore how the setup we have used could be further developed to operationalise compliance and enforcement of these new regulations across jurisdictions.

# Conclusion

Phase 1 of the CCIAO project has demonstrated the viability of PySyft to facilitate algorithmic auditing by external researchers. Four independent researchers successfully performed audits of recommender systems at LinkedIn and Dailymotion, whilst protecting the security and privacy of personal or commercially sensitive information through the combined application of remote data access and differential privacy.

These privacy guarantees, alongside the governance processes built into PySyft (namely the ability for the data owner to review and approve all code submissions), ensured that all relevant stakeholders at both LinkedIn and Dailymotion had sufficient confidence in the approach to make production data available through the platform.

We were able to carry out quantitative analysis of the recommender systems at both platforms to answer questions about how these algorithms shape the content recommended to users on the platform.

The successful outcomes from Phase 1 provide a strong foundation for future work on privacy-preserving third-party outcoming through CCIAO and, we hope, other initiatives.



# Acknowledgments

This project could not have been completed without the collective effort, expertise, and dedication of many individuals and teams. We would like to extend our gratitude to the following:

1. **LinkedIn:** Jon Adams, Tong Zhou, Adrian Rivera, Alex Murchison, Ryan Rogers, Rahul Tandra, Siyao Sun
2. **Dailymotion:** Sébastien Le Roux, Thomas Schmitt, Brice de la Brière, Cyrille Brun
3. **OpenMined:** Andrew Trask, Irina Bejan, Lacey Strahm, Stephen Gabriel, Zarreen Reza, Ishan Mishra, Osam Kyemenu-Sarash, Laura Ayre and the extremely talented engineering team that shipped and iterated on the platform that facilitated this project to match our research needs.

