

CrimRxiv

No Man's Hand: Artificial Intelligence Does Not Improve Police Report Writing Speed

**Ian T. Adams¹ Matt Barter² Kyle McLean³ Hunter Boehme¹
Irick A. Geary¹**

¹University of South Carolina, ²Manchester Police Department, ³Clemson University

Published on: Sep 11, 2024

DOI: <https://doi.org/10.21428/cb6ab371.4a6d42e9>

License: [Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License \(CC-BY-NC-ND 4.0\)](https://creativecommons.org/licenses/by-nc-nd/4.0/)

ABSTRACT

Objectives: This study examines the potential of artificial intelligence (AI) to reduce the time police officers spend writing reports, a task that consumes a significant portion of their workday.

Methods: In a pre-registered randomized controlled trial, we test this claim within the patrol division of a medium-sized police department (n=85), at the individual report level (n=755). Analyses utilize mixed-effects regression accounting for the nested structure of report-writing.

Results: AI assistance did not significantly affect the duration of writing police reports. Alternative specifications beyond those specified in the pre-registration, including a difference-in-differences approach observing report duration over a full year (n=6,084), confirms the null findings are robust.

Conclusions: Our findings contradict marketing expectations for the effect of this technology, suggesting no time-savings in report-writing can be expected when using AI-assisted report-writing. Several other potential effects remain possible and untested.

Introduction

The reduction of administrative burden through technological innovation has become a critical focus across both public and private sectors in recent years (Mergel et al., 2019). As organizations grapple with increasing demands and resource constraints, the pursuit of efficiency has led to widespread adoption of various technological solutions. Artificial Intelligence (AI) has emerged as a particularly promising tool in this endeavor, offering the potential to automate complex tasks, enhance decision-making processes, and significantly improve productivity (Brynjolfsson & McAfee, 2014). One of the most publicly available demonstrations of AI capabilities has been the rapid adoption of large language models (LLMs) that are able to produce human-like written products (e.g., ChatGPT) (Radford et al., 2019).

In the private sector, AI has been deployed in diverse applications, from streamlining customer service operations to optimizing supply chain management (Davenport & Ronanki, 2018). Companies like Amazon have leveraged AI to enhance warehouse efficiency, while financial institutions use AI-powered algorithms for fraud detection and risk assessment (Agrawal et al., 2017). Similarly, in healthcare, AI assists in diagnostic processes and administrative tasks, potentially reducing physician burnout and improving patient care (Alowais et al., 2023).

The public sector, while often slower to adopt new technologies, also recognizes the potential of AI to address long-standing efficiency challenges (Desouza, 2018). Government agencies are exploring AI applications in areas such as tax processing, benefits administration, and public transportation optimization (Mehr et al., 2017). These efforts aim to reduce paperwork, speed up service delivery, and ultimately improve citizen satisfaction with government services (Sun & Medaglia, 2019).

Just as private and public sector organizations are rapidly experimenting with AI and LLM tools to drive efficiency and output (Senadheera et al., 2024), the policing profession is historically adept at technology adoption to address its own concerns. In the realm of law enforcement, the adoption of AI presents both significant opportunities and unique challenges (Ferguson, 2017). Police staffing has become a critical issue in the United States, with agencies facing significant challenges in recruitment, retention, and retirements (Adams et al., 2023; Mourtgos et al., 2022) .

While some of this staffing distress is theorized to be related to acute reactions to significant social disruption beginning in 2020, there is good reason to believe that it is also related to persistent macro trends that are unlikely to markedly improve in the near future (Wilson & Heinonen, 2012). At the same time, call volumes reflect steady and growing demands for police services across the nation, and the primary predictor of police call response times is, unsurprisingly, the police labor available to meet that demand (Mourtgos et al., 2024). This situation underscores the need for law enforcement agencies to enhance their operational efficiency to keep pace with the increasing demand they encounter year after year (Wilson & Weiss, 2014).

Historically, technology has been the cornerstone of improving police efficiency (Stroshine, 2015). Major technological advancements have transformed policing practices and enhanced the capacity of law enforcement to manage their duties effectively. For instance, the introduction of motor vehicles revolutionized police transportation, enabling officers to respond to incidents more rapidly and cover larger areas (National Commission on Law Observance and Enforcement, 1931). Advancements in communication technology, such as the development of two-way radios, significantly improved coordination and response times, allowing for more efficient dispatching and information sharing (Leonard, 1938). The advent of computers and digital databases further streamlined administrative tasks and facilitated better data management, making information retrieval faster and more accurate. In recent years, analytically driven operational measures like hot-spot policing have leveraged data to identify and focus on high-crime areas, leading to more effective deployment of police resources and proactive crime prevention (Braga & Weisburd, 2022).

Building on the legacy of technological innovation, AI-assisted narrative generation represents the latest advancement with the potential to improve police report writing (Adams, 2024; Dement & Inglis, 2024; Ferguson, 2024). This technology is proposed to bring several benefits, including enhanced report quality, consistency, completeness, and efficiency in terms of reducing the time required for report writing (Axon, 2024). Given the increasing administrative burden on officers, reducing the time spent on paperwork is crucial to reallocating resources to more critical fieldwork. However, despite commercial claims that this technology will dramatically decrease the time officers spend manually writing initial reports (Keough, 2024), no experimental test of those claims has been reported to date. As is often the case, rapid adoption of police technology is often done in advance of the empirical record on the ability of the tool to achieve its aims, and avoid unintended consequences (Adams & Mastracci, 2019; Lum et al., 2017).

In this pre-registered randomized control trial, we focus specifically on the efficiency aspect of AI-assisted report writing, utilizing a commercial product from Axon called "Draft One." Our primary objective is to experimentally assess whether the use of AI tools can significantly reduce the time officers spend writing initial reports compared to traditional methods. By addressing this efficiency question, we aim to provide empirical evidence on the potential of AI technology to alleviate some of the operational pressures faced by modern police forces, particularly in an era of constrained staffing resources and increasing demands for accountability and transparency.

In both our pre-registered analysis and several alternative specifications, including a difference-in-differences analysis conducted over a full year, our findings consistently indicate that AI assistance did not significantly improve the speed of officers' report writing. While AI tools like "Draft One" may offer other benefits—such as improved consistency, accuracy, and report quality—the initial promises of this technology do not translate into the time savings that were anticipated.

The Promise of AI-Assisted Police Reports

AI-assisted narrative generation represents a key advancement with the potential to improve police report writing (Adams, 2024). This technology may bring several benefits, including enhanced report quality, consistency, completeness, and, efficiency in terms of reducing the time required for report writing (Lavezzorio, 2024; Ropek, 2024). Given the increasing imbalance between police staffing and public demand, reducing the time spent on paperwork is crucial to reallocating resources to more critical fieldwork. In other words, by reducing time spent on paperwork, police departments may be able to reduce police staffing woes as officers' time is freed up to answer more calls for service and spend more time in the community.

Previous scholarly work on AI-assisted police report writing is substantively non-existent. However, Ferguson (2024) engages in legal analysis about potential risks of the technology, including what he deems “generative suspicion.” Ferguson’s core critique is that traditional police report writing is such a critical part of the criminal justice system, that before we allow algorithms to affect the reports, we must better understand the first principles of police report writing. Failure to do so, and rush into adoption, risks a future where we “fundamentally reshape policing” with potentially negative consequences across the criminal justice system (Ferguson, 2024, p. 4).

In the absence of peer-reviewed findings necessary to establish competent priors regarding the purported effects, we must rely on information provided by the manufacturers of these commercial products. For instance, Axon, the world’s largest producer of body-worn cameras and conducted energy devices, recently introduced their "Draft One" product, which offers AI assistance for report writing. Axon’s press releases regarding Draft One quote an officer whose agency was testing the product as reporting that officers using the product spent 82% less time on report writing, and that the quality and completeness of their reports improved alongside the efficiency gains (Keough, 2024). These commercial claims are falsifiable, and in this study, we

focus specifically on the efficiency aspect of AI-assisted report writing. That is – does the use of AI tools significantly reduce the time officers spend on writing initial reports compared to traditional methods?

Method

Axon's Draft One is an AI-assisted report writing tool marketed as a solution to streamline the process of creating police incident reports (Keough, 2024). The system integrates with Axon body cameras, employing audio-to-text conversion technology to transcribe officer interactions. After an incident, officers can access the Draft One system, where they input basic incident details such as the type of crime, its severity, and arrest status. The system then generates an initial narrative draft based on the audio transcript and the officer-provided parameters.

The resulting narrative draft follows a standardized structure, typically including date, time, and officer identification, followed by sections detailing the incident's background, officer actions, suspect reactions, and the basis for suspicion or probable cause. Axon states that several features are incorporated to promote officer engagement and accuracy, including required information inserts, intentionally included errors for correction, and customizable thresholds for officer-generated content. The system reportedly requires officers to review, edit, and approve the final report, acknowledging its AI-generated origin and confirming its accuracy under oath.

At the core of Draft One's technology is ChatGPT 4, a large language model (LLM) developed by OpenAI (OpenAI et al., 2024).¹ LLMs are advanced artificial intelligence systems trained on vast amounts of text data, enabling them to understand and generate human-like text based on given prompts or instructions. In the context of Draft One, Axon creates transcripts from body-worn camera footage and then uses custom instructions to interact with the LLM API, requesting the generation of a police report based on the transcript and other provided parameters. Axons claims this technology allows for the rapid creation of structured, contextualized report narratives, and in turn saves time for the officer creating the report.

Agency Context

The study takes place within a medium-sized police department that agreed to participate in a pre-registered randomized controlled trial. The Manchester Police Department (MPD) engaged with the research team for an experimental trial to assess potential efficiency gains before full implementation. Manchester, New Hampshire, is a small city with an estimated population of 115,000, about 50 minutes north of Boston, in the New England region of the United States. The community is urban in nature and experiences crime and public safety issues consistent with other urban spaces in the country.

According to agency reporting, Manchester experienced a violent crime rate of 384 per 100,000 and a property crime rate of 1,960 per 100,000 in the calendar year 2023 (Aldenberg, 2023). MPD has primary law enforcement jurisdiction of the city, with an authorized strength of 271 full-time police officers and 67 non-

sworn personnel. Due to recruitment and retention challenges consistent with other large agencies (Adams et al., 2023), MPD's actual staffing consisted of 249 full-time officers and 54 non-sworn staff. The department is divided into six divisions, the largest being the Patrol Division, with a total staffing of 124 sworn officers, 106 of whom are patrol officers (who primarily respond to calls for service), and is overseen by a Captain (division commander), three lieutenants (shift commanders), and 14 sergeants (front-line supervisors).

During the study period, there were several noteworthy occurrences. First, as the study began, supervision within the patrol division changed. Each shift was assigned a new lieutenant (shift commander). These changes can disrupt the status quo in each shift. Additionally, late in the study several school resource officers (SROs) were added to the patrol division due to the end of the school year. These SROs were not included in the study. Lastly, an officer-involved shooting occurred in the last week of the study, which was a significant event for the department. The event drew significant resources and was labor-intensive for all involved.

Training and Implementation

Prior to participants using the Draft One tool, a structured training program was designed to familiarize officers with the new technology and study protocols. Initial communication was disseminated via email, providing participants with an overview of the technology and study objectives. Subsequently, in-person training sessions were conducted during patrol division roll calls from May 5 to May 12, 2024.

The training sessions were integrated into the existing organizational structure of daily roll calls, which typically serves as a platform for disseminating assignments and updates. This integration allowed for minimal disruption to normal operations while ensuring comprehensive coverage of the study population. Patrol supervisors were provided with a training roster to track participation and ensure all selected officers received the necessary instruction. The core of the training program consisted of a 17-minute instructional video, which participants viewed following their regular roll call duties. The video content was strategically designed to cover several key areas, including technology overview and functionality, departmental due diligence processes, operational integration with the agency's Records Management System (RMS), legal and procedural considerations, and best practices for optimal utilization of the Draft One tool.

The training curriculum emphasized three critical aspects of implementation. First, officers were instructed to initiate the incident report in the RMS prior to generating the narrative with Draft One, a crucial step for accurate timestamp tracking in data collection. Second, the importance of thorough review and verification of the AI-generated narratives was repeatedly stressed to ensure accuracy, completeness, and the removal of any erroneous or non-factual elements. Third, officers were trained in strategies to enhance the accuracy and detail of AI-generated reports, including techniques for clear verbalization of actions and observations during incidents, and providing comprehensive verbal summaries on Body-Worn Camera (BWC) recordings.

The training content was delivered through a webinar format, incorporating narrated screen recordings to provide visual guidance on the web interface usage. This multimedia approach was designed to accommodate

various learning styles and enhance retention of the operational procedures. Pre-study testing informed the training design, particularly the emphasis on verbalization techniques, which had been empirically shown to improve the accuracy and detail of generated reports. This evidence-based approach to training development underscores the iterative nature of the implementation process and the integration of preliminary findings into the study protocol.

Sample Characteristics and Randomization Procedure

The study sample comprised 85 police officers from the partner agency, representing a subset of the total patrol complement. This sample size is smaller than the full patrol division, and reflects various exclusions, including officers assigned to extended training programs, those on extended sick leave, military deployments, or administrative leave. Officers who opted out of the study were also excluded. Furthermore, newly hired officers still in the police academy or undergoing field training were not included in the sample.

Participants were randomly selected from the pool of willing officers and subsequently randomly assigned to either the control group ($n = 43$) or the experimental group ($n = 42$). The control group maintained their usual report writing procedures, while the experimental group received training on and utilized the AI-assisted narrative generation tool.

Table 1 presents the balance of key demographic and professional characteristics across the control and experimental groups. Randomization was done using the `randomizr` package in R (Coppock, 2023). Sample demographics largely align with national law enforcement workforce trends (Gardner & Scott, 2022), and the randomization process achieved successful balance across treatment groups.

The median age of participants was 31 years (IQR: 29.0, 34.0), with the control group slightly younger (median 30.0 years; IQR: 27.5, 33.0) than the AI group (median 33.0 years; IQR: 30.0, 36.8), though this difference was not statistically significant ($p = 0.5$). The median tenure was 3.50 years (IQR: 2.50, 6.20), with the AI group showing a marginally higher median tenure (4.20 years; IQR: 2.50, 7.48) compared to the control group (3.50 years; IQR: 2.25, 5.40), but again, this difference was not statistically significant ($p = 0.12$).

The sample was predominantly male (82%) and white (82%), reflecting broader trends in law enforcement demographics. The gender distribution was nearly identical across groups, with 81% male officers in the control group and 83% in the AI group ($p > 0.9$). Similarly, racial composition was balanced, with 79% white officers in the control group and 86% in the AI group ($p = 0.6$). Shift assignments were also relatively balanced ($p = 0.3$), with the largest proportion of officers working swing shifts (41%), followed by day shifts (33%) and midnight shifts (26%). The control group had a slightly higher proportion of officers on swing shifts (47% vs. 36%), while the AI group had more officers on day shifts (40% vs. 26%).

The balanced distribution across all measured variables, as evidenced by the non-significant p -values (all $p > 0.05$), indicate that the randomization process was successful in creating comparable treatment and control

groups. This balance strengthens the internal validity of the study, allowing for more robust causal inferences about the effect of the AI-assisted narrative generation tool on report writing outcomes.

Variable	Overall, N = 85 ¹	Control, N = 43 ¹	AI, N = 42 ¹	p-value ²
Age (years)	31.0 (29.0, 34.0)	30.0 (27.5, 33.0)	33.0 (30.0, 36.8)	0.5
Tenure (years)	3.50 (2.50, 6.20)	3.50 (2.25, 5.40)	4.20 (2.50, 7.48)	0.12
Sex				>0.9
Female	15 (18%)	8 (19%)	7 (17%)	
Male	70 (82%)	35 (81%)	35 (83%)	
Race				0.6
Non-White	15 (18%)	9 (21%)	6 (14%)	
White	70 (82%)	34 (79%)	36 (86%)	
Shift				0.3
Swings	35 (41%)	20 (47%)	15 (36%)	
Days	28 (33%)	11 (26%)	17 (40%)	
Midnights	22 (26%)	12 (28%)	10 (24%)	
¹ Median (IQR); n (%)				
² Pearson's Chi-squared test				

Table 1: Sample Statistics Balance Table

Sample Size & Statistical Power

We conducted a power analysis using the *pwr* package in R (Champely, 2020) to determine the required sample size for detecting a statistically significant difference in report writing time between the control and experimental groups. We used a two-sample t-test assuming equal variances, with a significance level (alpha) of 0.05 and a power of 80%. Based on historical data provided by the partner agency (mean report writing time = 54.63 minutes, standard deviation = 47.18 minutes), we estimated that 351 observations per group would be

needed to detect a relatively conservative effect size of 10 minutes reduction in report writing time for the experimental group. In the end, the study period included 755 observations (reports), and therefore the study is well-powered at the given metrics.

Data and Measures

Our sole outcome is report duration, observing the reports submitted by officers during the trial period (n=755). Our study drew upon the Manchester Police Department's (MPD) Records Management System (RMS). MPD utilizes Central Square's Enterprise RMS version 22.2.6. We used this audit reports from this system to create data on the time taken to complete incident reports and workstation usage for report completion. We extracted timestamps for report creation (when an officer opens a new template) and report submission (when an officer sends the report for review), along with unique workstation identifiers. This information allowed us to calculate the total (whole) minutes taken to complete each report.

Analysis

Data analysis for this study follows the pre-registered experimental protocol.² Pre-registered analyses are a preferred method for conducting experiments, such as the one presented here, as we state our hypotheses and the methods used to test the hypotheses prior to collecting data, thereby eliminating the possibility of p-hacking or other questionable research practices that artificially increase the likelihood of receiving a significant finding. Pre-registering experimental hypotheses has been shown to enhance the transparency and credibility of research by reducing bias and preventing data-driven modifications to hypotheses after results are known. This approach minimizes the risk of engaging in p-hacking or selective reporting, which can distort scientific findings. Studies have demonstrated that pre-registered experiments are less likely to report inflated effect sizes and more likely to produce replicable results, providing a stronger foundation for empirical evidence in fields such as criminology and psychology (Chin et al., 2023; Nosek et al., 2018, 2022). Consequently, pre-registration improves the overall rigor and trustworthiness of experimental research.

Given the nature of the data, where officers completed multiple reports over the study period, our pre-registration specifies a mixed-effects model to accommodate the repeated measures inherent in the data structure. This approach is suited to the hierarchical organization of the dataset—specifically, multiple reports nested within each officer and across various days. The mixed-effects model enabled us to control for individual variability between officers and consider the correlations between reports composed by the same officer.

The primary fixed effect in our model was the treatment variable, distinguishing between control and experimental groups. This distinction enabled us to estimate the average difference in report writing time attributable to the use of the AI tool. We incorporated a random intercept for each officer to recognize and model the natural variation in writing speeds—some officers are inherently faster or slower than others.

The general form of the mixed effects model we use is:

$$\begin{aligned} duration_i & \sim N(\alpha_{j[i]}, \sigma^2) \\ \alpha_j & \sim N\left(\gamma_0^\alpha + \gamma_1^\alpha(treatment), \sigma_{\alpha_j}^2\right), \text{ for id } j = 1, \dots, J \end{aligned}$$

Where:

- **duration mins_i** is the time it takes officer *j* to complete report *i*, measured in minutes.
- α_j is the average report writing time for officer *j*. This allows each officer to have their own baseline writing speed, recognizing natural variations in individual efficiency.
- γ_0^α is the average report writing time for the control group. This represents the baseline writing speed without the AI tool.
- γ_1^α is the effect of the treatment (using the AI tool) on report writing time. This coefficient will reveal whether the AI tool leads to a statistically significant difference in writing speed.
- **treatment** is a binary variable indicating whether the officer is in the control group (0) or the experimental group (1).
- σ^2 is the variance of the report writing times within officers.
- $\sigma_{\alpha_j}^2$ is the variance of the average report writing times between officers.
- *j* is the index for officers, ranging from 1 to *J* (total number of officers).
- *i* is the index for reports written by a specific officer.

Results

We proceed with the preregistered analysis using the mixed effect regression approach discussed above. The principal finding is that AI assistance did not significantly affect report writing duration. Results are reported in Table 2. Following the main results, to check the robustness of the finding, we provide four supplemental non-registered analyses, all of which support the main findings.

In the pre-registered protocol main model, treatment was associated with a non-significant reduction of report completion time, with wide confidence intervals ($b = -29.66$, $SE = 39.62$). Given the observed skewness in the outcome, the same model with a logged duration outcome measure was analyzed, confirming the non-significant effect of the AI assistance on report writing duration. Similarly, we evaluated a model that dropped the 5% longest reports, one that filtered to only reports less than four hours, and a final model with only reports less than one hour in duration. Across all specifications, treatment remained non-significant, demonstrating that AI assistance did not meaningfully impact report completion times regardless of the model used.

Our pre-registration also specified a supplemental test using a difference-in-differences model with fixed effects held by officer *id*, observing both control and treated officers' reports in the pre- and post-intervention period. Results for that specification, using one year of data on report duration ($n = 6,084$) were similarly statistically non-significant, and those results are reported in Appendix Table A1.

	Pre-Registration Protocol	Full Logged	95% Lower	< 240 mins	< 60 mins
(Intercept)	111.254***	3.755***	47.010***	50.641***	30.443***
	(27.070)	(0.072)	(2.467)	(2.929)	(1.162)
AI treatment	-29.657	-0.023	2.907	2.074	0.413
	(39.622)	(0.104)	(3.545)	(4.203)	(1.682)
SD (Officer)	44.389	0.321	10.843	13.142	3.722
SD (Observations)	518.811	0.938	31.164	36.335	15.221
Num. Obs.	755	755	717	732	496
R2 Marg.	0.001	0.000	0.002	0.001	0.000
R2 Cond.	0.008	0.105	0.110	0.116	0.057
AIC	11576.7	2113.2	7019.9	7393.4	4133.9
BIC	11595.2	2131.7	7038.2	7411.8	4150.8
ICC	0.0	0.1	0.1	0.1	0.1
RMSE	516.44	0.91	30.32	35.33	14.90
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001					

Table 2: Regression Results – Impact of AI Assistance on Report Writing Duration

Discussion

In a pre-registered protocol, we have provided the first experimental evidence on the impact of AI-assisted report writing technology on police officers' report writing efficiency. While there is widespread hope that efficiency gains could improve the ongoing staffing challenges faced by many agencies (Adams et al., 2023; Mourtgos et al., 2022), our results suggest caution.

The null effects observed in our study conflict with the broader literature on technological advancements improving efficiency in various sectors (Brynjolfsson & McAfee, 2014; Czarnitzki et al., 2023). The context of

policing, however, is known to be unique and researchers are warned to be context-sensitive when considering the potential effects of technology in the policing workplace (Koper et al., 2014).

Several potential policing realities may explain the null effects. Many officers, and indeed even many agencies, already utilize templates or other boilerplate prose for writing reports for common calls and offenses (Miller & Whitehead, 2014). To the extent that Draft One requires officers to fill in or confirm the detail of an incident, the process may not be substantially different from the template approach already commonly utilized (Adams, 2024). Another limit may be that a technology that assists in a report *narrative* does not substantively affect overall report duration. This is due to the realities of police report writing, in which the entire report engages the officer in more than just a narrative. For example, officers writing reports are typically required to complete a great deal of data entry, such as individual entries for every person they spoke to (complainants, victims, witnesses, and suspects), as well as any and all evidence or other property that came into the officer's possession during their response to the case (recovered property, drugs, weapons, etc.).

Therefore, even if AI technology like Draft One can streamline the narrative-writing process, it may not significantly reduce the total time required to complete a report. The bulk of police report writing involves meticulous data entry and documentation of various aspects of an incident that AI may not yet be equipped to manage efficiently. Furthermore, the rigid structures already in place, such as templates and standardized data fields, may limit the potential time savings from narrative assistance. These factors suggest that while technological advancements hold promise, their application in policing may face unique constraints that dampen their expected efficiency gains. Like other industries, artificial intelligence technology's impact on police productivity is context-dependent (Czarnitzki et al., 2023), and the complexities of law enforcement reporting present a distinct challenge that requires more tailored innovations to see substantial improvements in efficiency (Koper et al., 2014; Lum et al., 2017; Mastrobuoni, 2020).

Pushing Forward the AI-Report Writing Research Agenda

Our results should not be interpreted as a dismissal of all potential effects of AI-assisted report writing. Broadly, these effects can be categorized into efficiency, quality, consistency, and downstream consumption (Adams, 2024; Dement & Inglis, 2024). While our study found no significant time savings—contrary to the marketing claims surrounding AI—efficiency should not be the sole focus. Report quality remains a persistent concern in policing, with long-standing issues related to poor spelling, grammar, voice, and tone. AI assistance has the potential to address these issues, and Axon's internal study suggests that their Draft One system produces reports with improved terminology and coherence, while maintaining similar levels of completeness, neutrality, and objectivity (Axon, 2024). However, these findings require independent verification—as noted previously, Axon also claimed an 82% reduction in report-writing time. Future research should develop comprehensive metrics to evaluate the quality of AI-assisted reports, considering factors such as accuracy, completeness, and evidentiary value.

The consistency of report writing is another area where AI could play a beneficial role, potentially reducing variability between officers. However, this consistency might come at the cost of individuality and context-specific nuance, which are often crucial in police reports. Standardization could inadvertently lead to reports that are less reflective of unique incidents, potentially overlooking critical details that are important for legal proceedings or community relations.

Moreover, the downstream consumption of these reports—by courts, lawyers, community oversight bodies, media, and even academics—might also be impacted by the introduction of AI. AI-generated reports may be perceived as more uniform or polished, which could influence how they are interpreted or valued by different stakeholders. This could lead to positive outcomes, such as increased credibility and readability, but also negative consequences, such as a reduced sense of transparency or authenticity. There may be other “downstream” effects into the court system that emerge from AI-assisted report writing such as better evidence recording and quality report writing for prosecutors to charge and convict suspects (Boivin & Gendron, 2022). Consistently higher-quality AI-assisted reports might raise evidentiary standards in the criminal justice system, presenting challenges for cases based on traditionally written reports. More detailed reports could also require additional time for legal review, potentially creating new bottlenecks in the court system.

At the same time, the downstream effects of AI-assisted report writing on the criminal justice system may not be positive. Ferguson (2024) presents compelling concerns about AI-assisted police reports reshaping the criminal justice system. He argues that these reports could profoundly impact every stage of the process, from charging to sentencing. Prosecutors and judges may rely on AI-generated content for critical decisions without fully grasping its limitations or biases. Ferguson highlights potential discovery issues, questioning whether audit logs, prompts, and training data should be disclosed alongside the final report. At trial, he notes the challenges in cross-examining opaque AI-generated content. In plea bargaining and sentencing, especially for misdemeanors, these reports might disproportionately influence outcomes. Ultimately, Ferguson cautions that “generative suspicion” could erode human judgment and accountability in the justice system.

On the other hand, we should also consider the potential for agency-level efficiencies that may arise even when the initial report writing duration does not change, as observed in our experiment. If AI assistance improves the consistency and quality of reports, it is plausible that sergeants or supervisors responsible for reviewing and approving these reports may find fewer reasons to reject or require revisions. This could streamline the reporting process, reducing the time spent on back-and-forth edits and approvals, thereby enhancing overall efficiency within the agency. Moreover, the reduction in report rejections could allow officers to spend more time on patrol or other critical duties, further contributing to operational efficiency. Thus, while our findings suggest that AI assistance does not significantly reduce the time taken to write reports initially, its impact on the broader workflow and administrative processes within a police department could still offer valuable gains in efficiency. If this potential effect is realized, officers would spend less time revisiting and revising reports, increasing the operational time available for other duties (Chartrand & Verret, 2023).

Limitations

As with all experimental settings, our design emphasizes internal validity while acknowledging that external validity remains the burden of ongoing and future research. In other words, the primary limitation of our effort is its focus on a single agency. Replication studies across a variety of contexts are necessary and should include smaller and larger agencies, rural and metropolitan settings, and international contexts to validate and extend our findings.

Conclusion

We have provided the first experimental evidence of AI-assisted report writing in law enforcement, showing that despite vendor claims of 82% (Keough, 2024), real world testing resulted in no significant time savings. As we are entering a phase of police adoption of these tools, results should be interpreted cautiously. As seen in previous body-worn camera research, initial findings may not be consistently replicated across varied settings (Lum et al., 2019). Further research is needed to validate these results across diverse agencies and to assess long-term impacts on report quality, accuracy, and downstream criminal justice outcomes. Future studies should pay additional attention to potential unintended consequences and ethical considerations, particularly the effects on vulnerable populations and on core constitutional concerns (Adams & Mastracci, 2017; Ferguson, 2024).

The marketing narrative surrounding AI-assisted technologies has heavily emphasized time savings (Keough, 2024), but our experimental findings provide a strong challenge to this claim. As the inevitable tide of AI-assisted technologies comes to policing's shores, it is essential to approach the widespread adoption of AI technologies with a critical eye. As seen here, the promised efficiencies may not materialize as expected. Instead of assuming success, scholars and practitioners should be more open to the possibility that these tools might not deliver on all fronts and adjust our expectations accordingly.

Ethics

All participants were informed of the purpose of the study, the voluntary nature of their participation, and their right to withdraw at any time without penalty. They were assured of the confidentiality and anonymity of their responses. The University of South Carolina has provided IRB approval through Study # Pro00136198. This study was not funded, and the authors declare they have no conflicts of interest.

References

- Adams, I. T. (2024). Large Language Models and Artificial Intelligence for Police Report Writing. In *CrimRxiv*. <https://doi.org/10.21428/cb6ab371.779603ee>
- Adams, I. T., & Mastracci, S. H. (2017). Visibility is a Trap: The Ethics of Police Body-Worn Cameras and Control. *Administrative Theory and Praxis*, 39(4), 313–328. <https://doi.org/10.1080/10841806.2017.1381482>

- Adams, I. T., & Mastracci, S. H. (2019). Police Body-Worn Cameras: Effects on Officers' Burnout and Perceived Organizational Support. *Police Quarterly*, 22(1), 5–30. <https://doi.org/10.1177/1098611118783987>
- Adams, I. T., Mourtgos, S. M., & Nix, J. (2023). Turnover in Large Us Policing Agencies Following the George Floyd Protests. *Journal of Criminal Justice*, 88, 1–16. <https://doi.org/10.1016/j.jcrimjus.2023.102105>
- Agrawal, A., Gans, J., & Goldfarb, A. (2017). *What to expect from artificial intelligence*. MIT Sloan Management Review Cambridge, MA. <https://agrawal.ca/s/What-to-Expect-From-Artificial-Intelligence-b88l.pdf>
- Aldenberg, A. (2023). *2023 Annual Report: Manchester Police Department*. Manchester Police Department. https://www.manchesternh.gov/Portals/2/Departments/police/2023_Annual_Report.pdf
- Alowais, S. A., Alghamdi, S. S., Alsuhebany, N., Alqahtani, T., Alshaya, A. I., Almohareb, S. N., Aldairem, A., Alrashed, M., Bin Saleh, K., Badreldin, H. A., Al Yami, M. S., Al Harbi, S., & Albekairy, A. M. (2023). Revolutionizing healthcare: The role of artificial intelligence in clinical practice. *BMC Medical Education*, 23(1), 689. <https://doi.org/10.1186/s12909-023-04698-z>
- Axon. (2024). *Comparing quality between Officer-only and Draft One report narratives*. <https://www.axon.com/blog/examining-quality-and-bias>
- Boivin, R., & Gendron, A. (2022). An experimental study of the impact of body-worn cameras on police report writing. *Journal of Experimental Criminology*, 18(4), 747–764. <https://doi.org/10.1007/s11292-021-09469-8>
- Braga, A. A., & Weisburd, D. L. (2022). Does Hot Spots Policing Have Meaningful Impacts on Crime? Findings from An Alternative Approach to Estimating Effect Sizes from Place-Based Program Evaluations. *Journal of Quantitative Criminology*, 38(1), 1–22. <https://doi.org/10.1007/s10940-020-09481-7>
- Brynjolfsson, E., & McAfee, A. (2014). *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. W. W. Norton & Company.
- Champely, S. (2020). *pwr: Basic functions for power analysis* [Manual]. <https://CRAN.R-project.org/package=pwr>
- Chin, J. M., Pickett, J. T., Vazire, S., & Holcombe, A. O. (2023). Questionable Research Practices and Open Science in Quantitative Criminology. *Journal of Quantitative Criminology*, 39(1), 21–51. <https://doi.org/10.1007/s10940-021-09525-6>
- Coppock, A. (2023). *randomizr: Easy-to-use tools for common forms of random assignment and sampling* [Computer software]. <https://CRAN.R-project.org/package=randomizr>

Czarnitzki, D., Fernández, G. P., & Rammer, C. (2023). Artificial intelligence and firm-level productivity. *Journal of Economic Behavior & Organization*, 211, 188–205. <https://doi.org/10.1016/j.jebo.2023.05.008>

Davenport, T. H., & Ronanki, R. (2018). Artificial intelligence for the real world. *Harvard Business Review*, 96(1), 108–116.

Dement, C., & Inglis, M. (2024). Artificial intelligence-assisted criminal justice reporting: An exploratory study of benefits, concerns, and future directions. *Criminology & Criminal Justice*, 17488958241274296. <https://doi.org/10.1177/17488958241274296>

Desouza, K. (2018). Delivering artificial intelligence in government. *IBM for the Business of Government Report*.

Ferguson, A. G. (2017). *The Rise of Big Data Policing: Surveillance, Race, and the Future of Law Enforcement*. NYU Press.

Ferguson, A. G. (2024). *Ai-Assisted Police Reports and the Challenge of Generative Suspicion*.

Gardner, A. M., & Scott, K. M. (2022). *Census of State and Local Law Enforcement Agencies, 2018* (NCJ 302187). Bureau of Justice Statistics.

Keough, V. (2024, April 23). *Axon reimagines report writing with Draft One, a first-of-its-kind AI-powered force multiplier for public safety*. Axon IR. <https://investor.axon.com/2024-04-23-Axon-reimagines-report-writing-with-Draft-One,-a-first-of-its-kind-AI-powered-force-multiplier-for-public-safety>

Koper, C. S., Lum, C., & Willis, J. J. (2014). Optimizing the Use of Technology in Policing: Results and Implications from a Multi-Site Study of the Social, Organizational, and Behavioural Aspects of Implementing Police Technologies. *Policing: A Journal of Policy and Practice*, 8(2), 212–221. <https://doi.org/10.1093/police/pau015>

Lavezzorio, C. (2024, May 18). Fort Collins police testing artificial intelligence to speed up report writing time. *Denver 7 Colorado News (KMGH)*. <https://www.denver7.com/news/front-range/fort-collins/fort-collins-police-testing-artificial-intelligence-to-speed-up-report-writing-time>

Leonard, V. (1938). *Police communication systems*. University of California Press.

Lum, C., Koper, C. S., & Willis, J. (2017). Understanding the Limits of Technology's Impact on Police Effectiveness. *Police Quarterly*, 20(2), 135–163. <https://doi.org/10.1177/1098611116667279>

Lum, C., Stoltz, M., Koper, C. S., & Scherer, J. A. (2019). Research on body-worn cameras: What we know, what we need to know. *Criminology & Public Policy*, 18(1), 93–118.

Mastrobuoni, G. (2020). Crime is Terribly Revealing: Information Technology and Police Productivity. *The Review of Economic Studies*, 87(6), 2727–2753. <https://doi.org/10.1093/restud/rdaa009>

Mehr, H., Ash, H., & Fellow, D. (2017). Artificial intelligence for citizen services and government. *Ash Cent. Democr. Gov. Innov. Harvard Kennedy Sch.*, No. August, 1, 12.

Mergel, I., Edelmann, N., & Haug, N. (2019). Defining digital transformation: Results from expert interviews. *Government Information Quarterly*, 36(4), 101385. <https://doi.org/10.1016/j.giq.2019.06.002>

Miller, L., & Whitehead, J. (2014). *Report Writing for Criminal Justice Professionals* (5th ed.). Routledge. <https://doi.org/10.4324/9781315721354>

Mourtgos, S. M., Adams, I. T., & Nix, J. (2022). Elevated police turnover following the summer of George Floyd protests: A synthetic control study. *Criminology & Public Policy*, 21(1), 9–33. <https://doi.org/10.1111/1745-9133.12556>

Mourtgos, S. M., Adams, I. T., & Nix, J. (2024). Staffing Levels Are the Most Important Factor Influencing Police Response Times. *Policing: A Journal of Policy and Practice*, 18. <https://doi.org/10.1093/police/paae002>

National Commission on Law Observance and Enforcement. (1931). *The police*.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences*, 115(11), 2600–2606.

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, Robustness, and Reproducibility in Psychological Science. *Annual Review of Psychology*, 73(Volume 73, 2022), 719–748. <https://doi.org/10.1146/annurev-psych-020821-114157>

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., ... Zoph, B. (2024). *GPT-4 Technical Report* (arXiv:2303.08774). arXiv. <https://doi.org/10.48550/arXiv.2303.08774>

Radford, A., Wu, J., Clark, J., Askell, A., Lansky, D., Hernandez, D., Amodei, D., & Luan, D. (2019, February 14). Better Language Models and Their Implications [Technology]. *OpenAI*. <https://openai.com/blog/better-language-models/>

Ropek, L. (2024, April 23). Cops Are Now Using AI to Generate Police Reports. *Gizmodo*. <https://gizmodo.com/cops-are-now-using-ai-to-generate-police-reports-1851429617>

Senadheera, S., Yigitcanlar, T., Desouza, K. C., Mossberger, K., Corchado, J., Mehmood, R., Li, R. Y. M., & Cheong, P. H. (2024). Understanding Chatbot Adoption in Local Governments: A Review and Framework. *Journal of Urban Technology*, 0(0), 1–35. <https://doi.org/10.1080/10630732.2023.2297665>

Stroshine, M. S. (2015). Technological innovations in policing. In R. G. Dunham & G. P. Alpert (Eds.), *Critical issues in policing: Contemporary readings* (Vol. 911, p. 229). Waveland Press Long Grove, IL. [https://books.google.com/books?](https://books.google.com/books?hl=en&lr=&id=nqW3BgAAQBAJ&oi=fnd&pg=PA229&dq=Effectiveness+in+Policing+through+Computer+T)

[hl=en&lr=&id=nqW3BgAAQBAJ&oi=fnd&pg=PA229&dq=Effectiveness+in+Policing+through+Computer+T](https://books.google.com/books?hl=en&lr=&id=nqW3BgAAQBAJ&oi=fnd&pg=PA229&dq=Effectiveness+in+Policing+through+Computer+T)
[echnology&ots=CGfCPIBis8&sig=lhwZ9Ln_xFq9-OODd3Oap286Zko](https://books.google.com/books?hl=en&lr=&id=nqW3BgAAQBAJ&oi=fnd&pg=PA229&dq=Effectiveness+in+Policing+through+Computer+T)

Sun, T. Q., & Medaglia, R. (2019). Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, 36(2), 368–383. <https://doi.org/10.1016/j.giq.2018.09.008>

Wilson, J. M., & Weiss, A. (2014). Police Staffing Allocation and Managing Workload Demand: A Critical Assessment of Existing Practices1. *Policing: A Journal of Policy and Practice*, 8(2), 96–108. <https://doi.org/10.1093/police/pau002>

Appendix

Appendix Table 1: Difference-in-Differences Alternative Specifications

	Diff-in-Diff	DiD Logged	95% Lower	< 240 mins	< 60 mins
Pre-post	11.020	0.004	0.728	0.267	0.256
	(35.709)	(0.067)	(2.392)	(2.526)	(1.156)
Pre-post x Treatment	-30.962	-0.068	0.410	-0.447	-0.790
	(41.983)	(0.084)	(3.198)	(3.556)	(1.484)
Num. Obs.	6084	6084	5777	5908	4119
R2	0.011	0.090	0.091	0.096	0.103
R2 Adj.	-0.002	0.077	0.078	0.083	0.085
R2 Within	0.000	0.000	0.000	0.000	0.000
R2 Within Adj.	0.000	0.000	0.000	0.000	0.000

AIC	97226.1	16566.2	56525.3	59884.2	33861.9
BIC	97769.9	17109.9	57064.8	60425.6	34374.1
RMSE	704.90	0.93	31.79	37.92	14.47
Std. Errors	by: id	by: id	by: id	by: id	by: id
FE: id	X	X	X	X	X
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001					

Footnotes

1. In personal communications, Axon reports that the technical model currently in use is GPT4-mini, though tests are underway using later models of the LLM tool. [↪](#)
2. Pre-registration link blinded for review – registration provided as supplementary material for review. [↪](#)