



On March 25, 2024, at approximately 9:18 AM, I, Officer John Smith of the Local Police Department, initiated a stop on a silver Toyota with Indiana license plate number 151AD7 at the intersection of Seventh and South Street. Upon making contact with the driver, I identified myself as Officer Smith from the Local Police Department. I informed the driver that the reason for the stop was due to speeding in a school zone, where the speed limit is 25 miles per hour. [INSERT method of speed determination]. The driver acknowledged the infraction and expressed unawareness of the speed limit.

I requested the driver's license, which was provided without incident. [INSERT observation of driver's license condition or any anomalies]. I addressed the driver by the name on the license, Mike, and returned the license after a brief check. [INSERT any checks or verification performed by the officer].

Draft One

I issued a verbal warning to the driver for the speeding violation and advised him to slow down, especially in a school zone. [INSERT justification for issuing a warning instead of a citation]. The driver expressed appreciation for the warning and assured me that he would adhere to the speed limit in the future. [INSERT any final instructions or safety advice given to the driver].

The traffic stop concluded with the driver thanking me and both of us departing the scene without further incident. [INSERT any additional actions taken after the warning was given].

Study Overview

Comparing quality between Officer-only and Draft One report narratives

Table of Contents

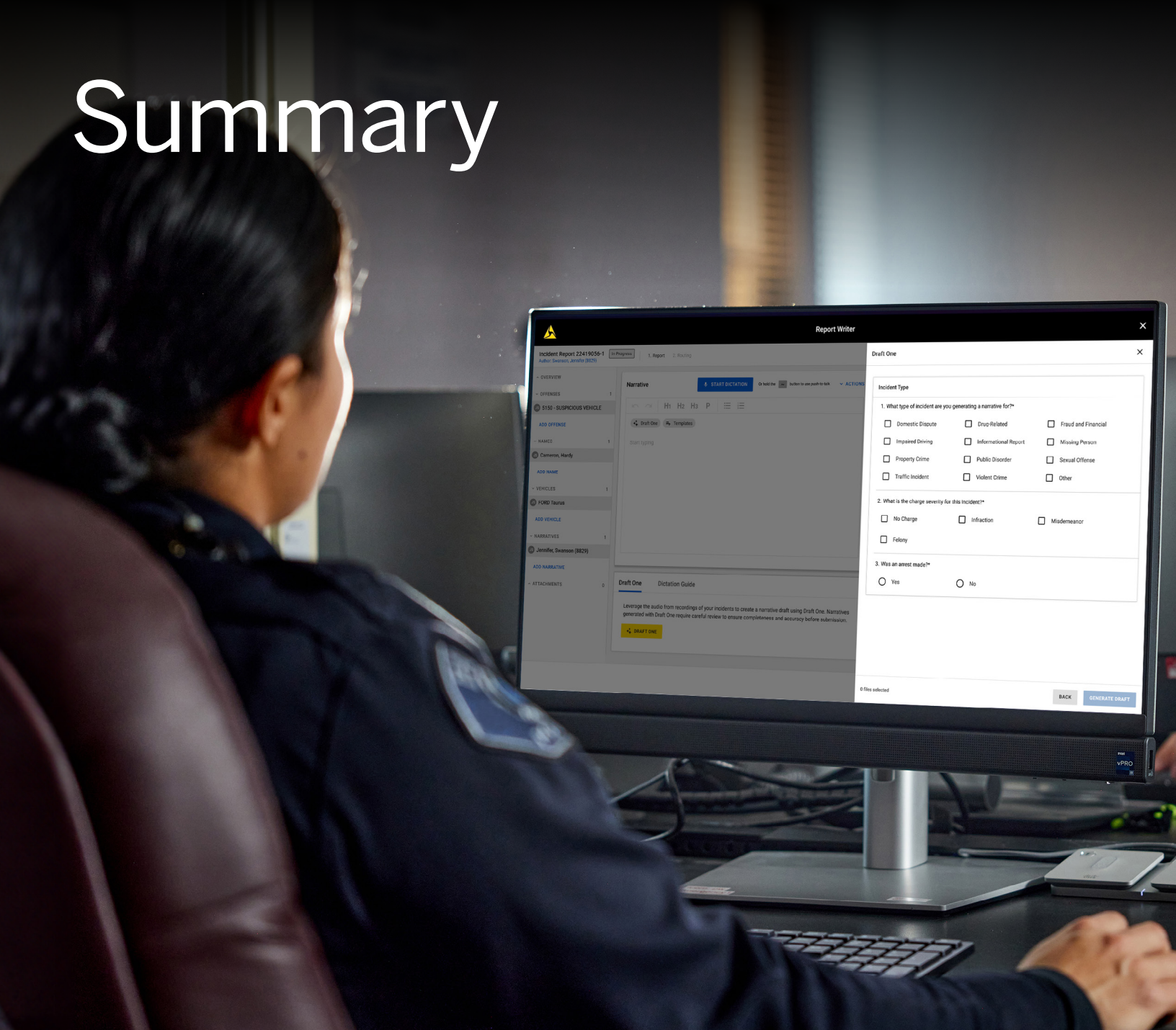
Summary

Methods

Results

Interpretation

Summary



Axon partnered with external experts to conduct a double-blind study to compare the effectiveness in quality between officer-only reports and reports created by an officer using Draft One. The study was conducted by twenty-four experts, based on their backgrounds in law enforcement, criminal law, and equity and inclusion.

This study demonstrates that Draft One has established a baseline for producing high-quality report narrative drafts. Results showed that Draft One narratives measured significantly better than officer-only narratives in the dimensions of terminology and coherence, and measure similarly to officer-only narratives in the dimensions of completeness, neutrality and objectivity. A significant factor that contributes to these results is the careful calibration done to Draft One to prevent speculation or embellishments which ensures that the generated narratives remain factual and reliable.

Methods



Study Design and Sample

This study employed a comparative analysis design to evaluate the quality of policing narratives written with and without artificial intelligence assistance. A total of 200 officer-only narratives were randomly sourced from a pool of reports generated within a three-month period. 80 narratives created with Draft One during the same time frame were also obtained. The selection criteria for narratives included a word count ranging from 150 to 1,200 words and the inclusion of specific types of incidents, which are defined below. This filtering process resulted in a balanced sample of 118 narratives, comprising 59 officer-only and 59 Draft One narratives, ensuring comparability between groups. Note: In this report, officer-only refers to report narratives written by an officer and Draft One refers to report narratives generated using AI but edited and submitted by an officer.

Narrative Selection

The narratives were further stratified based on incident types, which were predetermined as relevant to the study's aims. This stratification allowed for a focus on narratives of a similar nature and relevance. Narratives were also selected based on the overall word count for each narrative. The resulting range compiled a minimum of 150 words and a maximum of 1200 words. Axon selected report narratives from Axon's Customer Experience Improvement Program (ACEIP), the voluntary ethical and privacy-centric program that allows Axon to use data (e.g. Agency Content) stored in Axon Cloud Services to develop new products and improve our customers' product experience.

Expert Reviewers

Twenty-four experts independent of Axon were selected to review the narratives based on their backgrounds in law enforcement, criminal law, and equity and inclusion. The intention was to encompass a wide range of perspectives relevant to narrative quality assessment within the context of policing. Each expert was tasked with reviewing ten narratives, resulting in each narrative being evaluated twice to enhance the reliability of the assessments. Once the reviewer had access to the narratives, they were given seven days to complete the evaluation.

Blinding and Random Assignment

To minimize bias, experts were not informed whether a narrative was an officer-only report narrative or a Draft One report narrative. The assignment of narratives to experts was randomized to further control for selection bias and to ensure that each narrative was equally likely to be reviewed by any expert.

Rating Procedure

Experts were instructed to rate each narrative on a [Likert scale](#) from 1 to 5 across five domains: completeness, neutrality, objectivity, terminology, and coherence. These domains were chosen as they represent critical aspects of narrative quality in the context of policing reports. The Likert scale was defined as follows: 1 (Very Poor), 2 (Poor), 3 (Fair), 4 (Good) and 5 (Excellent). Definitions for each dimension can be found below:

- **Completeness** - All necessary elements required for a comprehensive understanding of the incident or crime appear in the narrative. Includes identification and roles of reporting persons, victims, witnesses, suspects, and other persons involved, if known. Includes specific details that contribute to a comprehensive understanding of the scene.
- **Neutrality** - All information is presented in a neutral and unbiased manner, avoiding any tone or toxic language that may suggest prejudice or signaling.
- **Objectivity** - The narrative does not include any opinions or subjective statements without supporting justification.
- **Terminology** - The language and technical terms employed in the narrative align with the expertise of the intended readers, which include members of the public, promoting appropriate communication and comprehension.
- **Coherence** - The narrative is presented in a coherent, logical and easily comprehensible way and utilizes proper grammar, spelling and professionalism.

Dimension	COMPLETENESS	NEUTRALITY	OBJECTIVITY	TERMINOLOGY	COHERENCE
Criteria	Does the narrative include facts relevant to the incident?	Is information presented in a neutral way?	Is the narrative free of personal opinions?	Is the language and terminology used appropriate for the intended use of report narratives?	Is the narrative written in a clear and understandable manner with proper formatting?
Definition	All necessary elements required for a comprehensive understanding of the incident or crime appear in the narrative. Includes specific details that contribute to a comprehensive understanding of the scene.	All information is presented in a neutral and unbiased manner, avoiding any tone or toxic language that may suggest prejudice or signaling.	The narrative does not include any opinions or subjective statements without supporting justification.	The language and technical terms employed in the narrative align with the expertise of the intended readers, which include members of the public, promoting appropriate communication and comprehension.	The narrative is presented in a coherent, logical and easily comprehensible way and utilizes proper grammar, spelling, and professionalism.
Excellent (5)	The narrative is fully complete, as outlined in the Definition.	The narrative is fully neutral, as outlined in the Definition.	The narrative is fully objective, as outlined in the Definition.	All terminology is appropriate, as outlined in the Definition.	The narrative is fully coherent, as outlined in the Definition.
Good (4)	The narrative is almost comprehensive, missing only 1 to 2 minor details that do not significantly affect the overall understanding of the incident.	The narrative is mostly neutral but contains 1 to 2 minor elements that could be interpreted as slightly biased. However, this does not significantly affect the report's credibility.	The narrative includes 1 to 2 minor subjective statements or opinions that do not significantly affect the overall objectivity.	The language and terminology are mostly appropriate , with only minor instances that might not perfectly align with the audience's understanding.	The narrative is mostly clear and coherent but may include 1 to 2 minor issues that don't significantly impede understanding.
Fair (3)	The narrative is mostly complete but lacks 3 to 4 elements and relevant descriptions that could be included to provide a more comprehensive understanding.	Generally, the narrative is neutral, but it has 3 to 4 elements that suggest a slight bias.	The draft includes 3 to 4 opinions or subjective statements but remains largely factual.	The language and terminology are generally appropriate but include some terms or phrases that might confuse the intended audience.	The narrative is clear but could benefit from revisions in 3 to 4 places to improve coherence and comprehension.
Poor (2)	The narrative is incomplete, 5 to 6 relevant descriptions are missing or are too vague, requiring substantial additions for a comprehensive understanding.	The narrative often appears biased, containing 5 to 6 elements that suggest a personal stance or prejudice. Significant revisions required.	The draft contains 5 to 6 opinions or subjective statements that significantly affect its objectivity, requiring substantial revisions.	The language and terminology are often inappropriate and do not align with the understanding of the intended audience, requiring substantial revisions.	The narrative is somewhat unclear and disorganized, requiring 5 to 6 revisions to improve coherence and understanding.
Very Poor (1)	The narrative has too few details, making it very difficult to understand the incident. Complete rewrite needed.	The narrative contains 7+ biases, which undermine the integrity of the report. Complete rewrite needed.	The narrative contains 7+ opinions and subjective statements, making it unreliable. Complete rewrite needed.	The language and terminology are completely inappropriate for the intended audience, making the report incomprehensible and requiring a complete rewrite.	The narrative is so unclear and disorganized that it is incomprehensible and requires a complete rewrite.

Word Choice Severity

This additional metric counts and classifies the number of descriptive words (verbs/adverbs/adjectives) used to refer to suspects in a narrative. The classification categorizes these words into following three categories based on the Severity of physical action contained in the word: Very Active, Active, Passive (example: "run": Very Active, "walk": Active, "sit": Passive). These word category counts are normalized by the total number of words classified. This is done to avoid penalizing narratives where more words are used to describe suspects. In this report, we mainly focus on the Very Active word count.

Results



Results: Officer-only

Officer-Only	Completeness		Neutrality		Objectivity		Terminology		Coherence		Total
Very Poor	4	3.40%	1	0.80%	1	0.80%	2	1.70%	7	5.90%	15
Poor	20	16.90%	13	11%	12	10.20%	11	9.30%	12	10.20%	68
Fair	14	11.90%	14	11.90%	15	12.70%	18	15.30%	20	16.90%	81
Good	43	36.40%	32	27.10%	35	29.70%	44	37.30%	40	33.90%	194
Excellent	37	31.40%	58	49.20%	55	46.60%	43	36.45	39	33.10%	232
	118	-	118	-	118	-	118	-	118	-	590

Results: Draft One

Officer-Only	Completeness		Neutrality		Objectivity		Terminology		Coherence		Total
Very Poor	4	3.40%	1	0.80%	1	0.80%	0	0%	0	0%	6
Poor	13	11%	10	8.50%	11	9.30%	4	3.40%	8	6.80%	46
Fair	25	21.20%	16	13.60%	15	12.70%	20	16.90%	25	21.20%	101
Good	34	28.80%	36	30.50%	37	31.40%	43	36.40%	37	31.40%	187
Excellent	42	35.60%	55	46.60%	54	45.80%	51	43.20%	48	40.70%	250
	118	-	118	-	118	-	118	-	118	-	590

Draft One / Officer-only Comparison

Hypothesis 1 (Group Equivalence): Draft One narratives measure at a level of quality equivalent to officer-only narratives in 99% of cases. (non-inferiority = 0.25)

Hypothesis 2 (Draft One Better): Draft One narratives measure at a level of quality better than officer-only narratives in 99% of cases.

Independent Narrative Ratings (N=113)

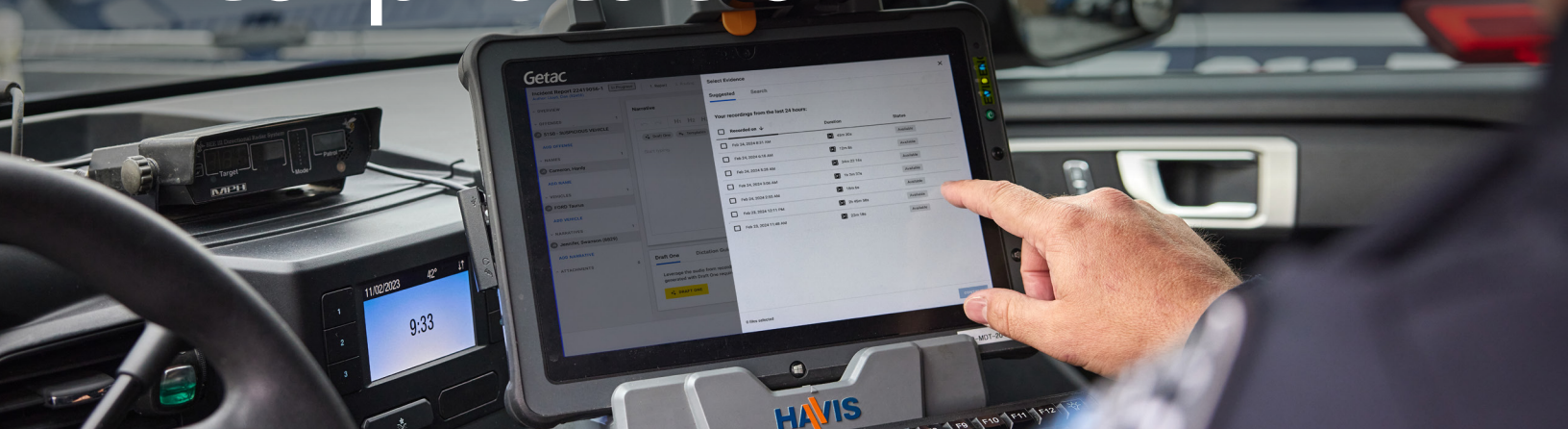
- **[Result]** In 3 of the 5 dimensions for quality — Completeness, Neutrality, Objectivity — Draft One measured as good as, or equivalent to, officer-only narratives.
- **[Result]** In 2 of the 5 dimensions for quality — Terminology and Coherence — Draft One measured significantly better than officer-only narratives.

Average scores for each category:	Overall	Completeness	Neutrality	Objectivity	Terminology	Coherence
Officer-Only (N=113)	3.93	3.73	4.12	4.11	3.97	3.75
Draft One (N=113)	4.06	3.83	4.13	4.12	4.20 (p=0.033)	4.05 (p=0.019)

Word Choice Severity

Officer-only (N=66)	1.9
Draft One (N=49)	0.76
Statistically significant?	Yes, p=0.007

Interpretation



Within the controlled parameters of this study, Draft One does significantly enhance the perceived quality of policing narratives in dimensions of terminology and coherence.

The team interprets this significant enhancement to quality in terminology as a function of Draft One, effectively employing appropriate language and technical terms on a more consistent basis regardless of narrative types. While narratives are primarily used as a tool by officers to summarize findings for detectives and prosecutors, this study signals the language used by Draft One to be more accessible to all audiences, based on our dimension of terminology.

This study also suggests using Draft One to assist in writing report narratives can significantly improve grammar, spelling and professionalism, contributing to overall coherence. The qualitative feedback from our expert reviewers and early access customers suggests officers often make grammar and spelling errors due to the high stress and time-sensitive nature of their job. This study suggests that Draft One enables officers to more reliably deliver coherent, flowing, professional narratives of events that occur in the field.

In the metric of word choice severity this study found Draft One to use significantly less very active words than Officer-only narratives when applying statistical analysis. This finding suggests that word choice in Draft One narratives may better support a subject's right to innocence until proven guilty. As a reminder, the final draft is always up to the officer. Users should review the drafts closely and adjust any word choices according to their best judgment.

This study demonstrates that Draft One has established a good baseline in producing high quality draft narratives. The Axon team will continually monitor the performance and effectiveness of the product, and will run a variety of studies to measure the impact of Draft One on report quality and larger the criminal justice process.

A previous version of this report stated that 60 narratives were included in both the control group and the experimental group. Upon further review, the correct number of narratives analyzed was 59 in each group.

As additional background, two narratives were excluded from the analysis after reviewer submission for methodological reasons. One officer-only narrative was excluded after it was discovered that only one of the three paragraphs intended for review had been presented to reviewers, rendering the narrative incomplete and unsuitable for evaluation. One AI-assisted officer narrative was excluded because it explicitly referenced the use of "Draft One," which compromised the blindness of the study.

As a result, the final analysis was conducted on 118 total narratives (59 control, 59 experimental), producing 1,180 total ratings. Any ratings associated with the excluded narratives were not included in the analysis. Their exclusion does not impact or alter the results or conclusions presented in the final study.



axon.com/draftone