

CASE STUDY

Newsroom: Brainerd Dispatch

About the newsroom

Brainerd Dispatch is a local newspaper serving the Brainerd, Minnesota area and is owned by Forum Communications. It is a mid-sized publication with a robust digital presence. Brainerd Dispatch provides comprehensive public safety coverage, including publication of police blotters, a summary of incidents responded to by law enforcement officers. Police blotters are received daily or weekly from a seven-county coverage area.

Project goals

The goal of the project was to automate the production of daily and weekly police blotters received via email. The newspaper proposed building an AI solution that can automatically assemble police blotter reports. The system would start with an input of police blotters received in PDF format via email. An additional goal was to create a searchable database of police incidents to power in-depth data journalism on trends in the area.

Why is this project important to the newsroom?

The Brainerd Dispatch's public safety coverage is highly valued by its readership as their most popular journalism category is local crime. The project aimed to maintain this coverage by automating brief- and explainer-writing while freeing up time for reporters to work on other tasks. By automating the production of police blotters, it could free up to 3 hours a day, up to 4 days for a week, and so more time could be spent building depth in crime reporting and adding more categories and trends to their coverage.

Engineering process

The AI system aimed to ingest police blotter reports in PDFs, extract newsworthy incidents, use the data from the report to populate a story template, and then use natural language generation to write a publishable story that went into the CMS.

The project was split into two distinct phases: the first was to automate the ingesting and processing of raw police blotters received via email; the second was to arrange the ingested data in a suitable story format.

The first phase was built by the University of Missouri while the second phase was built by commercial NLG vendor Lede AI. The selection of a commercial vendor for the automated writing was appropriate as it was unnecessary to rebuild an existing technology from scratch. Additionally, Lede AI was selected because the firm already provided automated high school sports reports to Forum Communications and had existing integrations into the Brightspot CMS.

Were the goals met?

The system met the goal of automatically crafting police blotter items as a basic story and delivering it into the CMS. However, the scope of the project was reduced during development. Initially, it was anticipated that the developers would be able to build out parsers to handle a variety of police blotter formats. But, the developed system handles blotters from three agencies. With the narrowed scope, the potential time savings from the project is reduced from the original goal of up to 3 hours a day. Additionally, there was no time left to build a method to export collected crime data to a spreadsheet to support data journalism.

Major challenges

A significant amount of time was spent on refining the parsers to increase their accuracy. Police blotter documents were presented in multiple formats, but a single parser still had to be able to ingest these different documents. It took longer to write logics that were able to distinguish and extract data from differently formatted documents. To address this, the team used the FITZ library and the PyPDF2 library to gain information about the structure of the PDF and extract data to directly upload it into the SQL Databases and then the User Interface. But the parsers only work for specific police blotter formats and two versions were ultimately developed. For other newsrooms to adopt the system, they would need to craft their own parsers to plug into the wider system.

Separately, the team at Lede AI developed new logic to handle public safety reporting data, which is not a standard commercial product. However, the Lede AI team ran into roadblocks and had to narrow their initial release to 15 categories of crime. Crime in other categories will have to be manually formatted by journalists.

STAKEHOLDER REACTION

“For years we have been manually going through Brainerd area police blotters and pulling information from them, formatting them and publishing them — a process that can be extremely time consuming. It is our hope that by using AI we can more quickly select and upload the information we wish to publish and save ourselves up to hours of time a week in the process.”

MATT ERICKSON
EDITOR

STAKEHOLDER REACTION

“Reporters are constantly tasked with doing more and more, but rarely are they given more time. This use of AI does just that. It gives them time back in their day to work on all the other aspects of their jobs, from adding video to taking photos, adding links and social media, to covering multiple beats and getting to more stories.”

RENEE RICHARDSON
MANAGING EDITOR

Future work

There is potentially significant value in building an all-purpose, custom trained, AI-driven parser for police blotters. This is neither a simple nor low-cost proposal, however, it would allow the system to operate using a single parser rather than requiring the crafting of multiple custom parsers.

Additional efficiencies could be gained by further exploring the capabilities of the open source usaddress NLP Python library.

Further, database management functions can be improved, including accomplishing the original goal of having a searchable database of incidents.

Link to repository

github.com/associatedpress/local-ai-brainerd-dispatch

Development team

First stage development was led by the University of Missouri under the guidance of Professor Prasad Calyam. Second stage development was done by Lede AI.

For Brainerd Dispatch: Matt Erickson, Renee Richardson

For Forum Communications: Rodney Holm, Steve Wagner, Mark Stahura, Jason Ingebrigtsen, Jason Graening, Tim Hanson

For Lede AI: Evan Ryan, Paula Stranges, Jothi Panchatcharam

For University of Missouri: Prasad Calyam, Hemanth Sai, Kiran Neupane, Roshan Neupane

For The Associated Press: Aimee Rinehart, Ernest Kung

Core components of the system

Input: Police blotter PDF file

Code: Python, HTML, JavaScript, CSS

Database: MySQL

Integrations: Lede AI via API, Brightspot via API

Output: Web portal, Lede AI, Brightspot CMS

Hosting: Amazon Web Services

Appendix

Figure 1. Process map

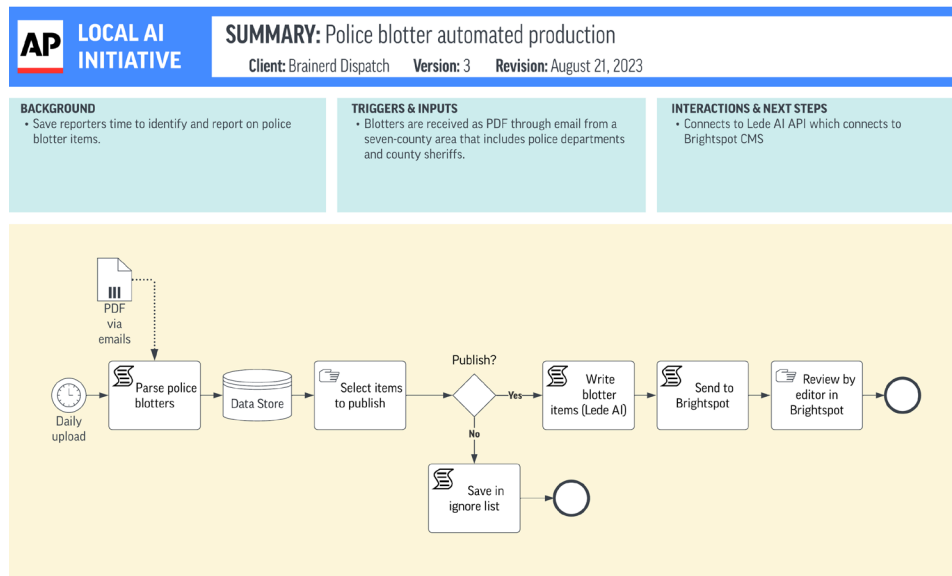


Figure 2. Database entity relationship diagram

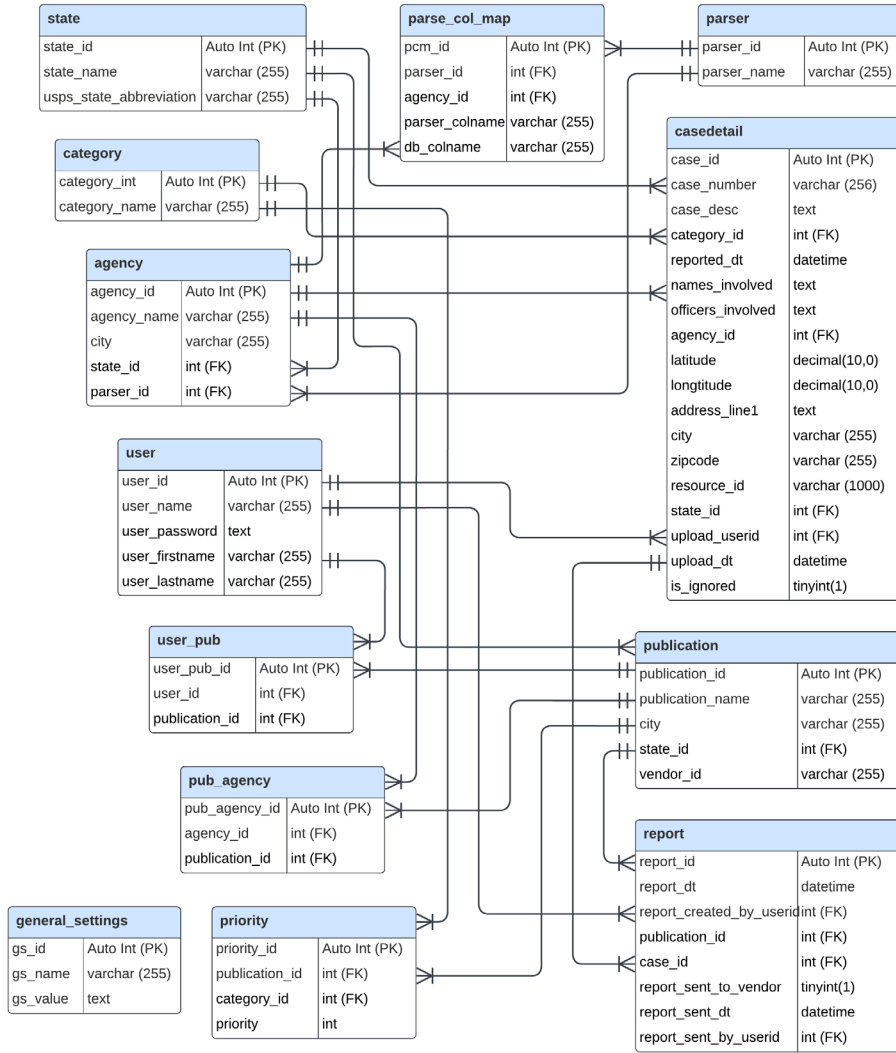


Figure 5. Brightspot CMS view of generated blotter story

Body

B *I* U | | | | | | | | | | |

ASSAULTS - A juvenile individual was arrested for misdemeanor 5th-degree assault at a group home for girls located in the 100 block of N 1st Street.

A citation was issued for misdemeanor 5th degree assault at the 100 block of 1st Street.

BURGLARY - A complaint was made, reporting the theft of a guitar in the 400 block.

THEFTS - A person was cited for misdemeanor theft-shoplifting at the 400 block of 6th Street.

A 21-year-old individual was arrested for misdemeanor theft in the 400 block of 8th Ave NE.

A citation was issued to an individual for shoplifting at the 600 block of Laurel Street.

A citation was issued for theft-shoplifting at the 500 block of B Street.

A theft was reported at Cash Wise Liquor - Brainerd in the 500 block of B Street.



THE ASSOCIATED PRESS

For more information, go to ap.org/ai