MERSEYSIDE POLICE

Machine Learning and Topic Modelling

(Understanding Demand and Crime MO)

S40(2) & **S40(2)**

Machine Learning and Topic Modelling

- Background & Context
- Understanding Calls Demand - Topic Modelling
- Automating MO

MERSEYSIDE POLICE

Topic Modelling – Background

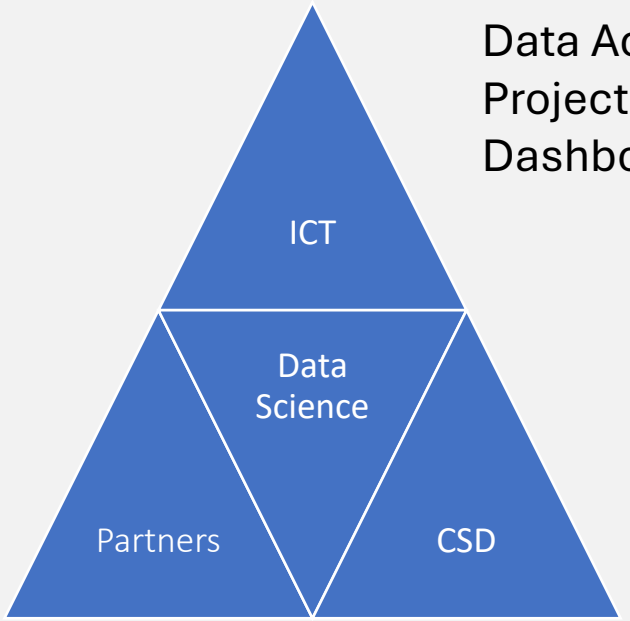
- Understanding our demand
- No record of non-incident calls
- Only audio files

...we don't know what half of our calls are about

MERSEYSIDE POLICE

Topic Modelling – Approach

- Merseyside Police not capable of full Data Science …
  …but does have skills!
- **Collaboration is key!!**



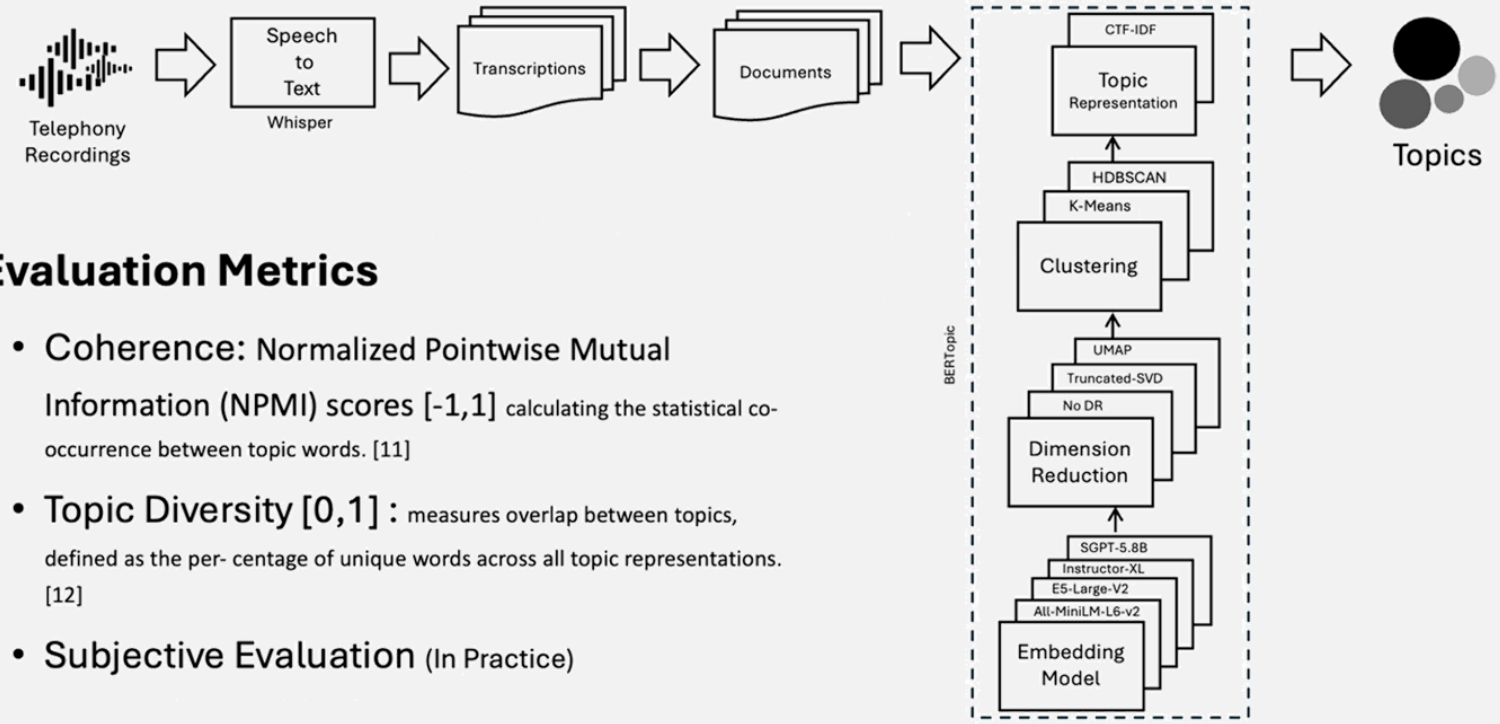
Data Acquisition/Processing
Project Mangement
Dashboards

Machine Learning (BU)
Data Pipelines (BS)
Output Provision (BS)

Analysis
Quality Assurance
Business benefits

MERSEYSIDE POLICE

Topic Modelling – Process



- **Evaluation Metrics**
  - Coherence: Normalized Pointwise Mutual Information (NPMI) scores [-1,1] calculating the statistical co-occurrence between topic words. [11]
  - Topic Diversity [0,1] : measures overlap between topics, defined as the per- centage of unique words across all topic representations. [12]
  - Subjective Evaluation (In Practice)

- Data Acquisition
- Transcription Issues
- Unnecessary topics
- Wrong topic labels
- Similar topic labels
- Numerous topics for a call

MERSEYSIDE POLICE

Topic Modelling – Example

Call audio file

Originally no insight into reason, or content, of call.

ML-generated Transcription
(speech-to-text)
393 words

Hello there, I'm calling from Whiston Hospital and one of the charge nurses. We've got two patients that are fighting outside. OK, bear with me one minute. Thank you. Whiston Hospital, yeah? Yeah. OK, bear with me. And have you got any weapons? Not that I'm aware of. We don't even know who the patients are. They've just pulled up in cars and two of them are fighting outside and causing a disturbance and they're blocking the entrance. We've rang security, but... Two males fighting, did you say? Yeah. Are we aware of outside the hospital? Literally outside A&E, which is our roundabout to the entrance. Outside A&E? Yeah, outside the A&E department, yeah. I can still hear shouting from there while I'm on the phone too, as well. Any descriptions? There's quite a few people. There's five or six people at the moment. There's mainly two males, but I've just come straight back in to ring you. No worries, it's fine. Bear with me a minute. We've got a sense of an emergency. Thank you. Sorry, security have been notified, but you're unsure if they're there? Yeah, I think they're coming now, but... OK, they are. And the males are blocking, is that right? Yeah, they've been blocking the entrance because they've got the car arguing and... The entrance to A&E? Yeah, the roundabout to A&E where all the patients pull up the drop-off area. OK, it's the drop-off area. No, this is just like, sorry, delaying it if I'm OK. Have you got a description of the white males? There are two white males, yeah. I haven't got any other descriptions of that, though. I think they'll all be on camera, as well, if needed. So if you just speak to security they'll all be on camera. And we don't know any of their names or anything? No, not yet, no. I can try and find out. I'll do some more information before you come and try and find out. Or why they're fighting. Yeah. All right, no worries. What's your contact number? [redacted] yeah. [redacted] All right, log number 1000 of today's date and we have dispatch on an emergency response. We've got two controls on read. All right, perfect. Thank you very much. All right, thank you. Cheers, thank you. Cheers, bye-bye. Thank you.

'Whisper' performance:
93% accurate

LLM-produced Summary
(reduced token length)
53 words

Two patients are fighting outside A&E at Whiston Hospital, blocking the entrance. Security has been notified but their presence is uncertain. The fight involves two white males, and there may be additional individuals involved. The hospital has logged the incident (log number 1000) and has two controls on standby for an emergency response.

'LLama2' performance:
98% accurate

LLM-generated Topic Label
(single topic per call)

Violence

Topic label accuracy currently being tested

MERSEYSIDE POLICE

Topic Modelling – Top 10 Topics

Emergency calls
73,578

Non-Emergency calls
53,065

Enquiry calls
49,235

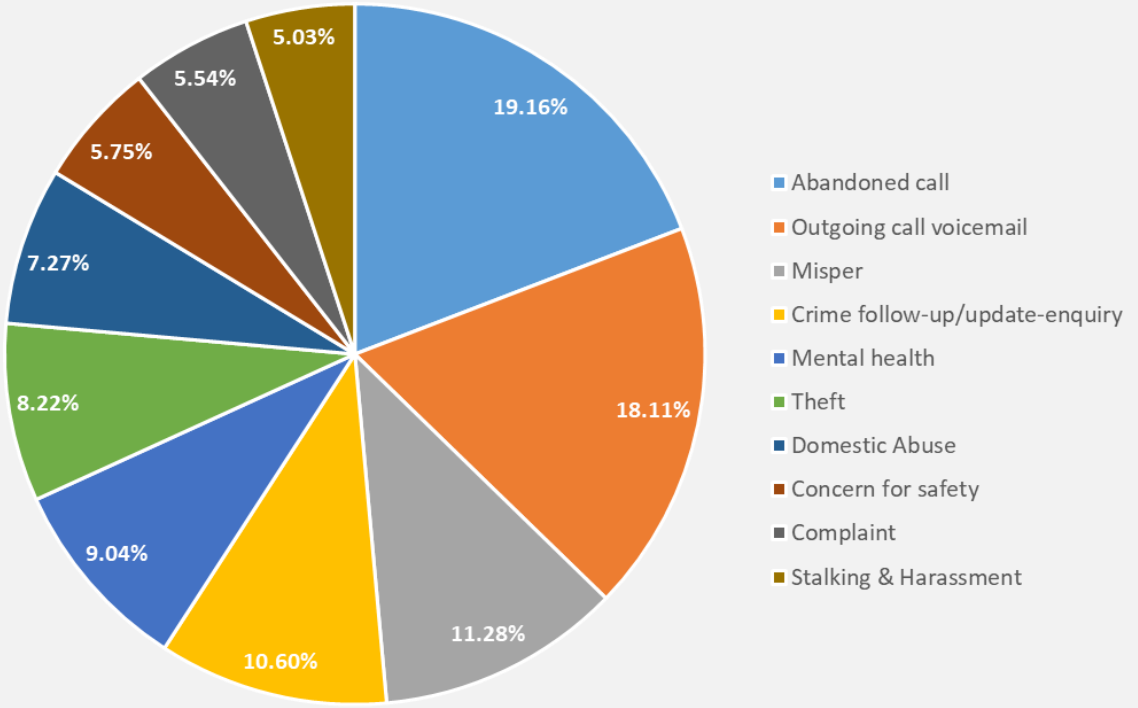
Apr-24 to Jun-24

ML clusters calls into
82 topic groups
(282 distinct topic labels within this)

E.g.,

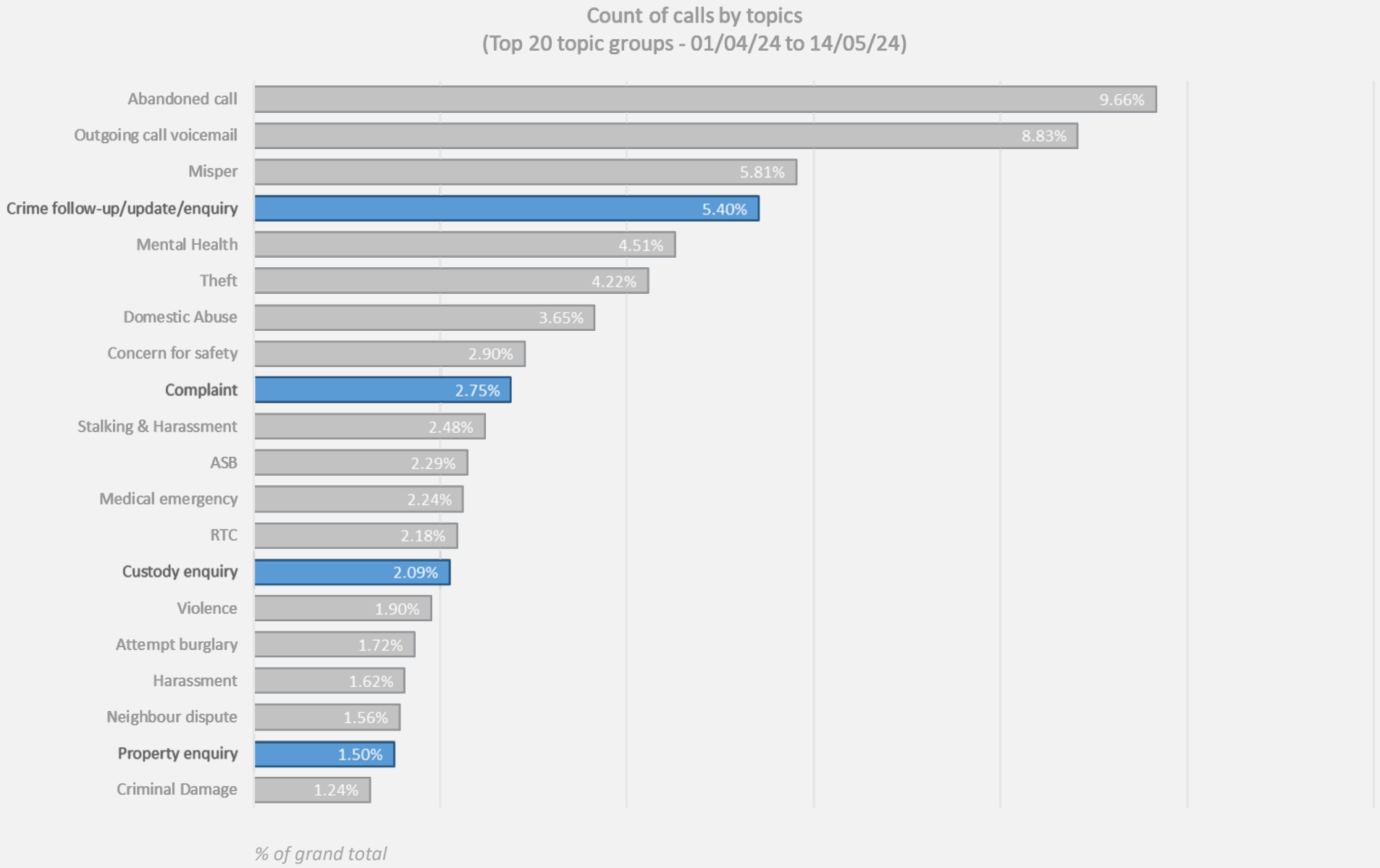
Topic Group	**Topic Labels**
Concern for safety:	38_Elderly Wanderers 57_Teenagers Jumping Into Water 73_Elderly Stranded 150_Welfare Check 280_Safeguarding Incident

Top 10 call topic groups



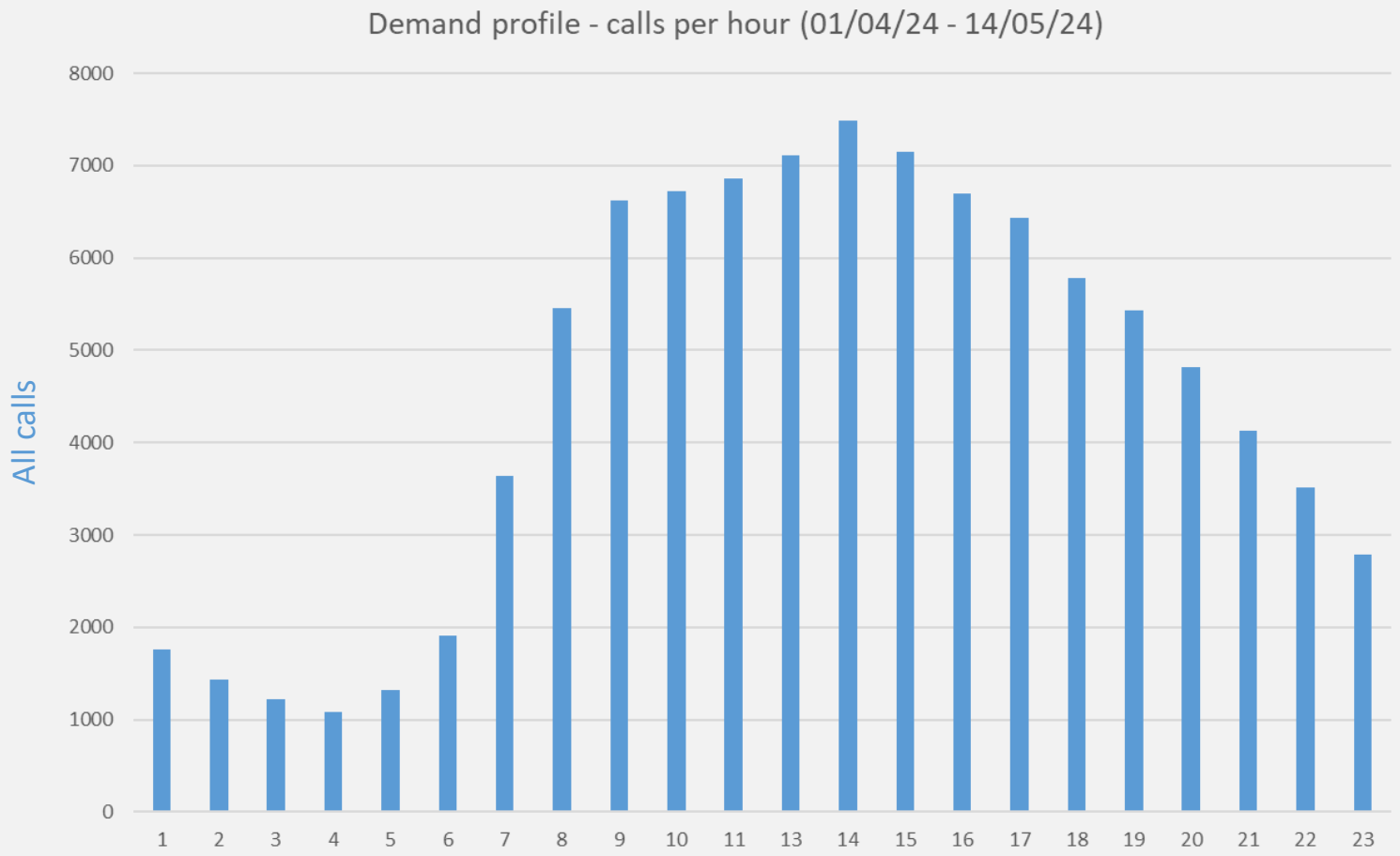
MERSEYSIDE POLICE

Topic Modelling – Top 20 Topics

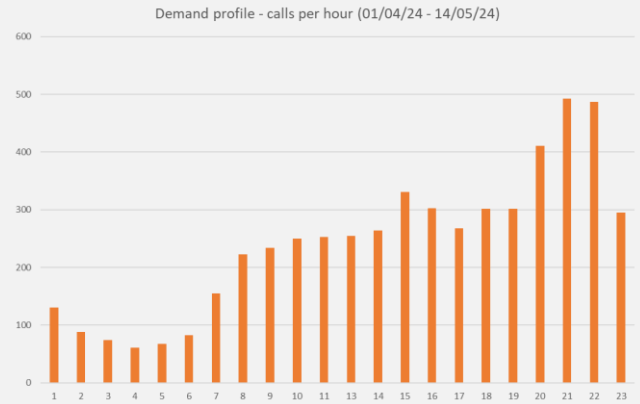


MERSEYSIDE POLICE

Topic Modelling – Demand Profile

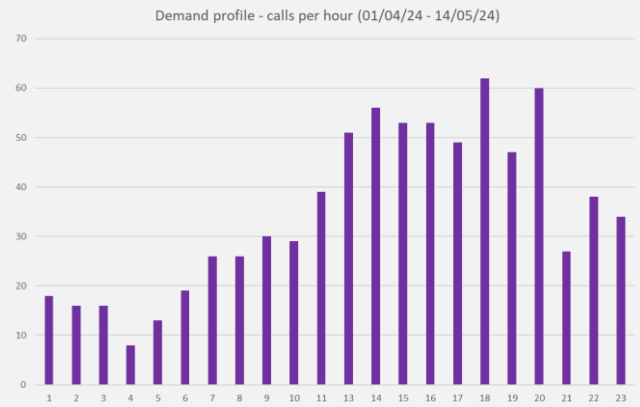


Misper



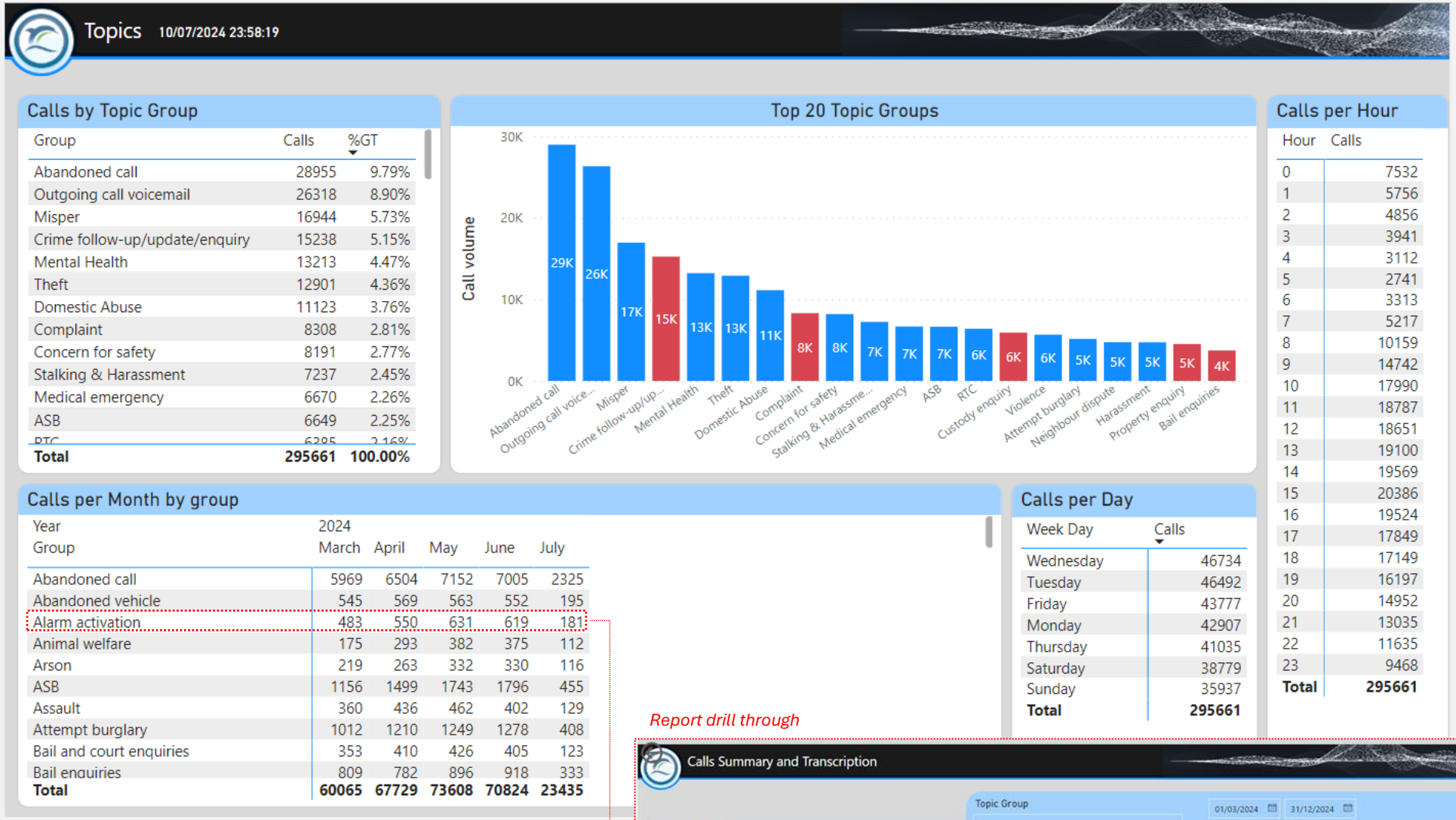
Crime follow-up/update/enquiry

Chase up deployment

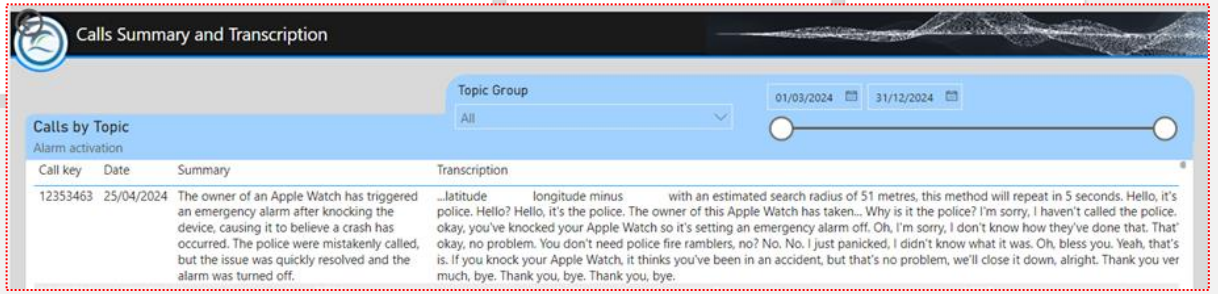


MERSEYSIDE POLICE

Topic Modelling – Dashboard



Report drill through



MERSEYSIDE POLICE

Topic Modelling – Next Steps

- Analysis of data to inform strategic insight
- Ongoing work to further test, and improve, accuracy
- Additional developments of MI dashboard
- Technical enhancements to process
- Working group in place to consider how data can be utilised further (e.g. sentiment analysis)

MERSEYSIDE POLICE

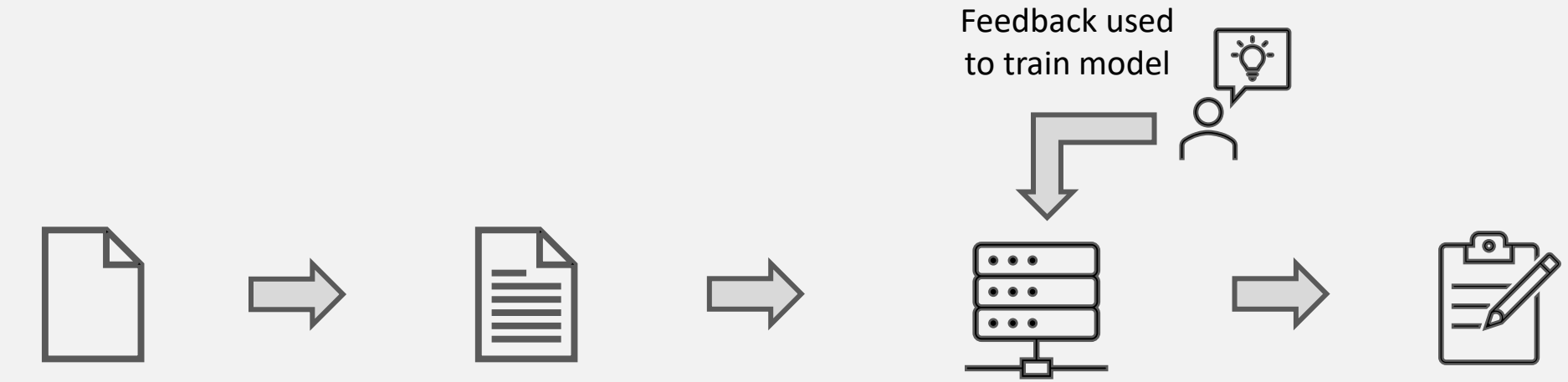
Automatic MO – Background

- Unsatisfactory Data Quality
- Find and Fix process

- ...existing 'Get it right second time' process is inefficient

MERSEYSIDE POLICE

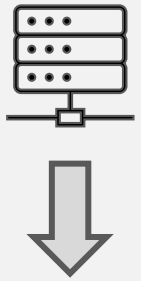
Automated MO – Process



Automated MO – Example

Input

Offender has entered rear garden of property and climbed up to flat roof outside window when he has been seen by victim's daughter who has returned home. Suspect has then shouted "RUN" and fled.



Output

1. The location of the burglary was **Dwelling**.
2. The actual point of entry was **Window - upper rear**.
3. The entry method was **Use code**. This is inferred because the offender opened the window, implying they used a code to unlock it.
4. Damage was caused to the **Rear flat roof door**.
5. **Unknown**. The occupation of the offender is not mentioned in the provided input.
6. Summary: A person attempted to burgle a dwelling by entering through the rear window on the upper floor. They were spotted by the victim's daughter, who returned home unexpectedly, causing the offender to shout "RUN" and flee the scene. The CSI was called to examine the scene, and no obvious forensic opportunities were found, although some objects were noted to be out of place. The offender's description was obtained from the witness statement, and further enquiries are required.

MERSEYSIDE POLICE

Automated MO – Examples -Structured and Free text MO

Withheld in full S31 S38

Automated Structured MO – Accuracy

Burglary

MO Field / Information	% completed in Niche *	Mixtral accuracy
Location	99%	63%
Entry Point	41%	88%
Entry Method	37%	87%
Attempted Entry Points	22%	75%
Damage Caused	22%	78%
Offender Occupation	21%	33%
Free-text Summary	100%	87% - 99%

** Apr-24 to Jun-24*

Sexual Offences

MO Field / Information	% completed in Niche *	Mixtral accuracy
Location	90%	59%
Victim/Offender Relationship	97%	79%
Victim Activities	45%	75%
Act of Violence	6%	77%
Sexual Acts	44%	87%
Offender Behaviour	6%	88%
Offender flee scene	3%	50%
How Victim/Offender Met	N/A	60%
Offender Occupation	38%	48%
Free-text Summary	98%	89% - 100%

** Apr-24 to Jun-24*

Violence

Currently being tested

MERSEYSIDE POLICE

Automated MO – Next steps

- Outputs to be aggregated into groups
- Further testing and accuracy improvements
- MI reports/dashboards
- Analysis of key themes (Intelligence and Strategic Analysis)
- Apply to broader offences

MERSEYSIDE POLICE