



# ModelOps for enhanced decision-making and governance in emergency control rooms

Kay Lefevre<sup>1</sup> · Chetan Arora<sup>1</sup> · Kevin Lee<sup>1</sup> · Arkady Zaslavsky<sup>1</sup> · Mohamed Reda Bouadjenek<sup>1</sup> · Ali Hassani<sup>1</sup> · Imran Razzak<sup>1</sup>

Accepted: 2 April 2022  
© The Author(s) 2022

## Abstract

For mission critical (MC) applications such as bushfire emergency management systems (EMS), understanding the current situation as a disaster unfolds is critical to saving lives, infrastructure and the environment. Incident control-room operators manage complex information and systems, especially with the emergence of Big Data. They are increasingly making decisions supported by artificial intelligence (AI) and machine learning (ML) tools for data analysis, prediction and decision-making. As the volume, speed and complexity of information increases due to more frequent fire events, greater availability of myriad IoT sensors, smart devices, satellite data and burgeoning use of social media, the advances in AI and ML that help to manage Big Data and support decision-making are increasingly perceived as “Black Box”. This paper aims to scope the requirements for bushfire EMS to improve Big Data management and governance of AI/ML. An analysis of ModelOps technology, used increasingly in the commercial sector, is undertaken to determine what components might be fit-for-purpose. The result is a novel set of ModelOps features, EMS requirements and an EMS-ModelOps framework that resolves more than 75% of issues whilst being sufficiently generic to apply to other types of mission-critical applications.

**Keywords** ModelOps · MLOps · Emergency management · Data governance · Requirements engineering

## 1 Introduction

Worldwide, reports indicate more intense fires amid unusually high temperatures and dry conditions across the Siberian tundra (Kutukov 2021), Arctic Russia (Deacon 2020),

Alaska, Canada (Berger 2021b), California (Berger 2021c), Oregon (Berger 2021a) and Greece (Baltas 2021). Climate change is very likely to continue to make weather more extreme and wildfires (bushfires, forest fires) more frequent and destructive (Field et al 2012).

The increased fire danger globally means shorter intervals between fire, increased intensity, fewer fires extinguished and faster spreading events (Parry et al 2007). For example, the frequency of very high and extreme fire danger days in south-east Australia is expected to rise by up to 70% by 2050 (Hennessy et al 2005). As the length of the fire season extends, the window of opportunity for fuel reduction burning contracts further into winter (Parry et al 2007).

The severity and frequency of bushfire events has worsened as the impacts of human-induced climate change are felt (Field et al 2012). Bush fires also have a large financial impact, with the 2009 Black Saturday bushfires in Victoria, Australia conservatively costing \$AU4.4billion (Teague et al 2010). The trend is for this to get worse, with emergency response leaders reporting that the severity and length of the bushfire season has continued to increase (Peacock et al 2021).

---

✉ Chetan Arora  
chetan.arora@deakin.edu.au

Kay Lefevre  
lefevrek@deakin.edu.au

Kevin Lee  
kevin.lee@deakin.edu.au

Arkady Zaslavsky  
arkady.zaslavsky@deakin.edu.au

Mohamed Reda Bouadjenek  
reda.bouadjenek@deakin.edu.au

Ali Hassani  
ali.hassani@deakin.edu.au

Imran Razzak  
imran.razzak@deakin.edu.au

<sup>1</sup> School of Information Technology, Deakin University, Geelong, VIC, Australia

Incident control-room operators, whether human or machine, are making decisions based on multiple sources of heterogeneous, unstructured data that continues to grow in frequency, volume and complexity (Shahrah and Al-Mashari 2017; Sun et al 2020). The AI/ML tools that support incident response are also increasingly complex (Zhao 2021), making it more difficult for operators to understand or audit the tools reliability.

This paper aims to investigate the area of emergency management systems (EMS) and develop high-level requirements that focus on improved data management and AI/ML governance. Specifically, we investigate the role of ModelOps in EMS. ModelOps (short for model operations) is a holistic approach to automating the deployment, monitoring, governance and continuous improvement of analytics models, so that they can quickly progress from the lab to production (Brethenoux et al 2018). ModelOps in EMS will enable reduced incident control-room response times, more transparent decision-making and overall reduced risk.

**Contributions.** The contributions of this paper are i) a review of governance technologies suitable for EMS; ii) an analysis of ModelOps as a framework to support EMS; iii) the extension of ModelOps and creation of requirements to map it to EMS; and iv) the evaluation of the proposed EMS-ModelOps framework, as a novel, generic and portable framework for supporting EMS.

**Structure.** This remainder of the paper is structured as follows. Section 2 contextualises the findings from the literature review. Section 3 describes ModelOps technologies, proposes a framework for combining technologies and a novel ModelOps feature list. Section 4 develops a novel set of EMS requirements and performs a gap analysis against the ModelOps feature list. Section 5 proposes methods to evaluate the resulting “EMS-ModelOps framework”. Section 6 presents threats to the validity of the design approach and results. Finally, Sect. 7 concludes the work and suggests further areas for research.

## 2 Literature review

Research indicates that the problem of managing huge volumes of heterogeneous data such as continuously streamed sensor time-series (Dugdale et al 2021), images (Asif et al 2021), satellite data (Routray et al 2019; Gozzard 2021; Kua et al 2021) and social media data (Thomas et al 2019) would benefit from improvements in data discovery and management technologies (Barika et al 2019; Sun et al 2020). In addition, it is likely that both humans and machines in the EMS incident control-room would operate more effectively if they had access to categorised (Kachaoui et al 2020), contextualised (Hassani et al 2018), indexed information and

tools that support good governance (Hummer et al 2019; Afyouni et al 2020):

- AI/ML processes that assign context to IoT data (Hassani et al 2018) or use ontologies like OWL (Kachaoui et al 2020) can enable other AI/ML tools to more easily discover, collect and analyse heterogeneous Big Data (Dugdale et al 2021).
- Continuous spatio-temporal indexing of social media data (Afyouni et al 2020) and knowledge storage (Buntain et al 2020) provides an opportunity to visualise the evolution of a disaster in near-real-time and provides a source of test and training data for disaster ML algorithms (Alam et al 2018; Thomas et al 2019).
- Recent improvements to algorithms for low-resolution image processing from geostationary satellites has delivered more frequent assessments of fire events, operating both night and day (Engel et al 2021).
- Satellite-laser data transmission, many thousands of times faster than radio transmission speeds, is likely to enable growth in receipt and processing of real-time fire-event images by 2024 (Gozzard 2021).

Compounding the issues of Big Data discovery, receipt and management, research indicates that the AI/ML tools that support incident response are increasingly complex, lack transparency and are not reproducible (Zhao 2021):

- Machine-enabled decision-making is perceived as increasingly black box (Moraffah et al 2020) as AI/ML becomes more complex (Zhao 2021).
- Operational teams need to understand ML/AI recommendations to have confidence in decision-making (Raglin et al 2021). The European Union (Commission European 2021) has proposed that machine-enabled decision-making systems must have human oversight and traceability.
- Emerging approaches to support human-machine decision-making are to quantify the Uncertainty of Information (UoI) (Raglin et al 2021) or apply eXplainable AI methods (XAI) (Arrieta et al 2020).
- Applying DevOps practices to Machine Learning algorithm development and modelling can improve operations for real-world applications (Karamitsos et al 2020).
- Increased computations and multiple ML models requires an infrastructure and platform for end-to-end lifecycle management, e.g. data processing, data validation, model design, model training, model evaluation, quality checks, deployment and maintenance (Zhou et al 2020).

In summary, research indicates that it is too complex for EMS incident control-room teams to manage an ever-increasing volume of unstructured data (Sun et al 2020),

to understand what machines (ML/AI) are doing (Moraffah et al 2020) and therefore have confidence in their contribution to decision-making (Commission European 2021). The advent of more complex Big Data (Zhao 2021), more frequent disaster events (Field et al 2012) and increasing use of AI/ML for discovery, management, modelling and governance of data is perceived as a risk - it is a “Black box”, as shown in Fig. 1.

The ModelOps approach (Brethenoux et al 2018) to managing datasets and modelling artefacts during AI/ML processes is a recent addition to business enterprise tools (Gartner 2021). (DataRobot, a commercial supplier, has an extensive explanation of MLOps for Machine Learning operations on their website (DataRobot 2021)). ModelOps builds on traditional DevOps best practices (Karamitsos et al 2020), addressing issues of artefact version control, data and model provenance and quality control checks by governing the end-to-end lifecycle of AI/ML processes (Hummer et al 2019; Zhou et al 2020). Preliminary analysis indicates that “XOps” technologies (a collective term for the suite of DataOps, ModelOps, MLOps, DevOps, etc. (Gartner 2021)) could improve EMS incident control-room decision-making by streamlining the management of heterogeneous Big Data and complex AI/ML processes.

### 3 What is ModelOps?

#### 3.1 Background

The need for AI model governance in the corporate sector was an active area of research by Gartner prior to 2018 (Brethenoux et al 2018). Based on corporate surveys (Gartner 2021) and ongoing sector analysis (Feinberg and Thanaraj 2020; Sicular and Vashisth 2020; Vashisth et al 2020b), Gartner identified an increasing need to operationalize AI modelling to quickly generate business value. DataOps (Feinberg and Thanaraj 2020) principles for data governance and DevOps (Hummer et al 2019) principles for implementation could be integrated with AI modelling processes, with rapid deployment and re-useability as key business goals. The term ModelOps (Brethenoux et al 2018)

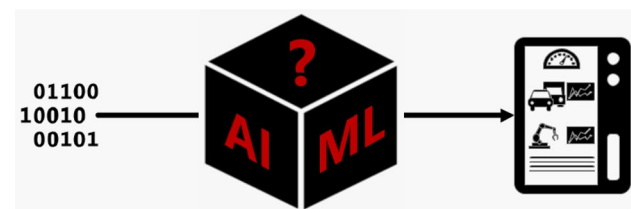


Fig. 1 Artificial intelligence and machine learning tools to create information are increasingly perceived as “Black box” processes

describes AI model operationalization; but in this report the proposed framework includes a small number of features from DataOps (Vashisth et al 2020b) that should improve EMS operations.

#### 3.2 Proposed ModelOps framework

Fig. 2 shows the proposed ModelOps framework, including data pipeline/workflow features and applies DevOps principles.

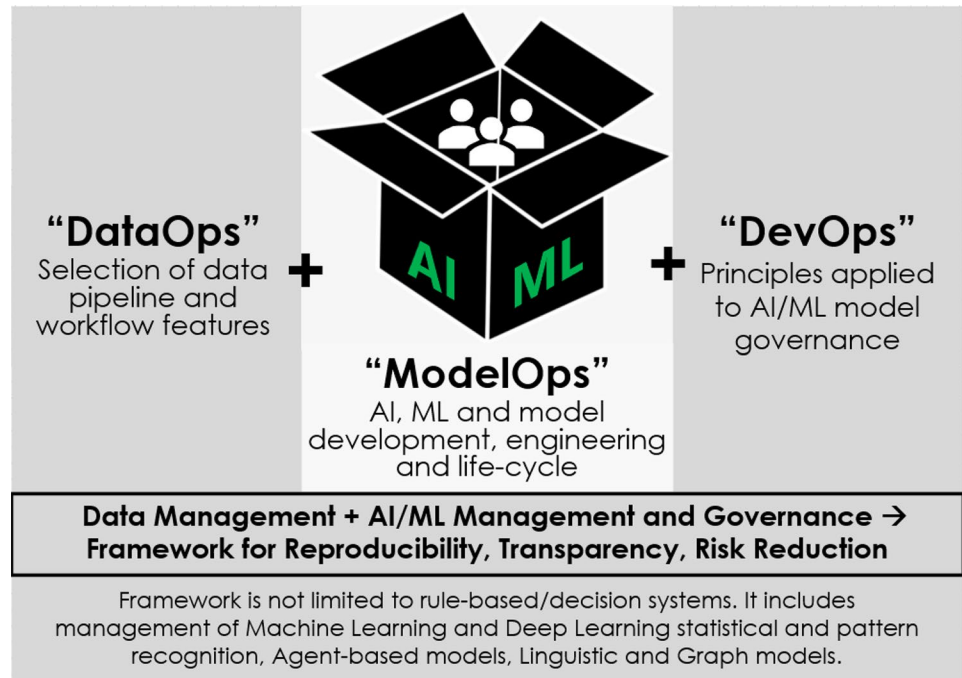
The proposed framework describes the governance and lifecycle management of data and AI models, but the models are not limited to rule-based/decision systems (Brethenoux et al 2018; Choudhary et al 2020). They include Machine Learning and Deep Learning statistical and pattern recognition models (Vashisth et al 2020a), Agent-based models, Linguistic models (e.g. speech recognition, predictive text, machine translation, sentiment analysis) and Graph models (Choudhary et al 2020). (Graphs provide linked contexts, showing the relationships among entities using Graph-shaped data. Beyer (Beyer 2020) uses the term “Active Metadata” to describe metadata that enables Graphing. Hummer et.al. (Hummer et al 2019) propose a ModelOps architecture that includes a Graph database).

Gartner surveys the corporate sector annually (Gartner 2021) and undertakes Information Technology research. In 2018 (Brethenoux et al 2018) the top-rated ModelOps goals for business domain experts were (i) able to assess the quality of AI models in production, (ii) promotion or demotion of AI models without a full dependency on data scientists or ML engineers and (iii) connecting model metrics to business KPI’s. In 2020, Forbes (Wu 2020a) interviewed Chief AI Architect of ModelOp.com, Stu Bailey. In his experience with commercial customers, “...a properly operating model can dramatically change the topline performance of a particular business unit. Integration between the business and compliance is critical”. In 2021, ModelOp.com surveyed 100 AI-focused executives regards the status of scaling up model operationalization (Bailey et al 2021). Survey respondents reported (i) an average of 270 models in production, (ii) low satisfaction ratings for their capacity to operationalise models and (iii) 80% reported difficulty with model compliance.

#### 3.3 ModelOps feature list

In 2020, ModelOp.com released a request for proposal (RFP) template to enable organisations to assess their own ModelOps functional requirements (ModelOp.com 2020). This is used to help develop 17 conceptual features in Table 1 below. Many of the original ModelOp.com requirements are too technology-centric for this analysis; for example the requirement for ModelOps software to be “Implementation Agnostic” describes its

**Fig. 2** ModelOps framework: a combination of DataOps, ModelOps and DevOps



coverage of model types, languages, execution environments and execution locations. Whilst these are important features they are outside the scope of high-level, conceptual framework development.

Table 1 below includes a mixture of DataOps, ModelOps and DevOps features that are relevant to our framework in Fig. 2 above. The first three features, F1–F3, under “Data Pipeline Management”, represent a subset of features usually associated with DataOps (Feinberg and Thanaraj 2020). The last feature, F17, relates to DevOps. The majority of features relate to AI/ML management and governance. For example, feature F12 “Dependencies, provenance, auditability” manages all dependencies used to execute a version of a model, including the management of dataset provenance, both inputs and outputs. The aim of creating provenance records is to improve queries and auditability.

The recommended feature list is a high-level synthesis of ideas from commercial resources such as ModelOp.com (ModelOp.com 2020), DataRobot (DataRobot 2021), DataBricks (Zumar and Uhlenhuth 2021), Microsoft Azure (Hanly and Sekar 2021) and analysis of research by Garter (Brethenoux et al 2018; Vashisth et al 2020a, b; Feinberg and Thanaraj 2020; Sicular and Vashisth 2020; Beyer 2020) and Forbes (Wu 2020a, b). In the next section these features are mapped to the EMS requirements.

## 4 EMS requirements development and gap analysis

### 4.1 Background

Shahrah and Al-Mashari (2017) investigate information systems and features to support emergency response. They report a lack of widespread use of AI/ML in EMS, with the exception of intelligent capabilities for decision-making within “Expert Systems”. The adoption of “Agent-Based Simulations” and “Case-Based Reasoning” tools are inhibited by poor response times and integration issues. In their 2017 survey, there is limited reference to tools for streamlining AI/ML processes, model-algorithm pipelines, managing multiple model scenarios, automated decision-support tools such as bots, artefact version control, provenance of inputs and outputs, or enabling human verification of machine processes.

Sun et al (2020) research the application of AI techniques to process disaster-related data. They identify 26 AI methods and 17 application areas, across four disaster phases. Of the research challenges identified, data management and AI governance issues include:

**Table 1** DataOps, ModelOps, DevOps features, F1–F17

| Category                       | Feature   | Description  |
|--------------------------------|---|--|
| Data pipeline management       | F1. Standard dataset metadata, taxonomies, federation | A standard dataset definition with active metadata, e.g. OWL or standard ontologies for IoT devices. Enables data discoverability, graph enablement and federation. Used for dataset versions, inputs and outputs  |
| Data pipeline management       | F2. Data orchestration                                | Taking data from multiple sources, combining and organising it, making it available for analysis. A supply of quality-coded, contextualised data   |
| Data pipeline management       | F3. Datasets for training and testing                 | Managing training and testing datasets. Need high-quality, large datasets to train and test AI. Users can manipulate the datasets, enabling them to better understand AI decisions and set KPIs. KPIs can be tested with real-world data   |
| AI governance                  | F4. Standard model metadata                           | A standard model definition with active metadata, which can be customised  |
| AI governance                  | F5. Ingest and standardise new AI                     | Standard import process for new AI models/algorithms. Reproduce AI published elsewhere. Process to test models against the Standard Model Definition and release for use   |
| AI governance                  | F6. Manage rules, compliance, performance and alerts  | Define and manage the rules, decision tables or keyword tables for compliance and performance through a simple interface, e.g. case-based reasoning (CBR) chatbots/workflows, thresholds for statistical performance, accuracy. Set custom / business KPIs, error messages. Set the triggers for alerts and the settings for notification management |
| AI governance                  | F7. Manage access controls and approvals              | Role-based and group-based access to datasets, models, workflows, rules. Manage who can query and command the system. Manage the assigned authority to approve workflow steps and alternative approvers  |
| AI governance                  | F8. Transparent querying and decision-making          | Manage the lifecycle of decision-making. Record user queries and commands (decisions). Manage approvals, for transparent and auditable human or machine decisions  |
| AI governance                  | F9. Audit report                                      | Generate an audit report that includes all steps in the model execution, with metadata, components and results   |
| Re-useability, reproducibility | F10. Inventory / storage and code repository          | Enable viewing and management of models and all of their component parts, e.g. model code, coefficients, weights, etc. Integrate with source code managers like GitHub   |
| Re-useability, reproducibility | F11. Manage artefact documentation                    | Dataset, model and model artefact documentation, version-controlled, auditable. The knowledge about a model can be packaged and transferred, making it easier to understand and re-use   |
| Re-useability, reproducibility | F12. Dependencies, provenance, auditability           | Manage all dependencies that are used to execute a version of a model, including the test/metrics. Manage dataset provenance (input and output). These records can enable historical queries and auditability  |
| Re-useability, reproducibility | F13. Model version control                            | Take a snapshot of a model in time, including all of the model's source code, artefacts, documentation, metadata and results   |
| Re-useability, reproducibility | F14. Reproduce a model run                            | Reproduce a model run, enabling user's to step through the process in a simple interface. If the same model produces different results, why?   |
| Re-useability, reproducibility | F15. Champion/Challenger testing                      | Compare a "champion" model or ensemble of models against a proposed "challenger", using various statistical and technical performance metrics e.g. Uncertainty of Information ranking  |

**Table 1** (continued)

| Category                       | Feature  | Description   |
|--------------------------------|--|---|
| Re-useability, reproducibility | F16. Standard Console, Dashboard                 | A console for model management and visualisations. Use approved templates or create custom reports for metrics, KPIs, scorecards  |
| DevOps                         | F17. Iteration and endpoint management, rollback | Apply DevOps practices and tools to model iterations, rollout, rollback. The promotion or demotion of AI should be possible without a full dependency on data scientists or engineers |

- Lack of enough quality data in some areas for accurate predictions using AI models, or data are incomplete due to changing disaster situations.
- The ability of AI to process, manage, and learn data decreases as data volume and complexity increases; especially within a reasonable response timeframe.
- EMS operational teams require user-friendly AI tools with interfaces that require minimal technical expertise.
- Results from AI processes should be repeatable, with improved interpretability and explainability. The replicability of predictive outputs improve AI’s trustworthiness, by capturing processes, data and parameters.

### 4.2 EMS requirements

Figure 3 presents the results of the investigation into EMS issues. The analysis has been guided by existing theory, recent developments in research and the project review of multiple EMS evaluations (Li et al 2017; Tsai et al 2019; Basak et al 2020; Damacharla et al 2020). As per project scope, the high-level requirements focus on data management, AI/ML management and governance functions.

The requirements are a novel framework (Hummer et al 2019; Sun et al 2020) for addressing problems managing huge volumes of heterogeneous Big Data and complex “black box” AI/ML, through the identification of gaps and the application of ModelOps features. Similar work to operationalise AI/ML is emerging in the commercial sector but

| Collect   | Analyse   | Display   |
|---|---|---|
| <p>R1: Discover, categorize and ingest increasing volumes of semi- and unstructured, multi-modal Big Data (IoT, social media)</p> <p>R2: Discover, categorize and ingest increasing volumes of satellite fire-event data</p> <p>R3: Create spatio-temporal and image datasets from unstructured social media data</p> <p>R6: Discover additional datasets where known volumes are insufficient to make valid predictions</p>  | <p>R4: Automate and rapidly classify information from phone/mobile calls</p> <p>R5: Effectively search very large spatio-temporal datasets and historical data</p> <p>R12: Analyze social media to track feelings and reactions of the public; geo-spatial sentiment mining using natural language processing</p> <p>R13: Detect and manage false alarms</p> <p>R14: Re-configure and test existing workflows or simulations during an emergency event and re-deploy solution</p> | <p>R11: Results from AI models should be explainable, repeatable, replicable</p> <p>R16: User-friendly AI interfaces that require minimal technical expertise for practical use</p> <p>R17: Real-time support for tools that require a high level of competence in deployment</p> |
| <p>R7: Version control of files and other artefacts to enable visualization of the history of changes</p> <p>R9: Capture the processes, data, and parameters for experiments to become repeatable</p>   | <p>R8: Logging/tracking of decision-making activities. These may be interactive, iterative and involve new data, criteria or goals.</p> <p>R10: Record and enable audit of complex, collaborative decision making</p>   | <p>R15: Test new fire-event algorithms or models as they are published</p>  |
| <p><b>Based on the literature review, most of the above type of functions are required for improved data management and AI/ML governance in EMS, but they are not implemented. Some functions are managed manually or are semi-automated (e.g. R1, R2, R4, R8, R10). According to Sun et al. (2020), the key issues are i) lack of enough quality data for accurate predictions using AI, ii) increasing data volumes and complexity, iii) lack of user-friendly AI tools and iv) inability to repeat, interpret or explain AI processes.</b></p> |   |   |

**Fig. 3** High-level requirements, grouped by collect, analyse, display phases

bushfire emergency management systems are uniquely complex due to their mission-criticality, i.e. the risk to human life (Commission European 2021) and need for real-time information.

For simplicity, Fig. 3 presents 17 requirements by phase (Collect, Analyse, Display). For example, R1 is “*Discover, categorize and ingest increasing volumes of semi- and unstructured, multi-modal Big Data (IoT, social media).*” This means that an EMS should enable humans and/or machines to do this function. R1 is assigned to “Collect” because data discovery and ingest logically occurs before data is analysed or results displayed. Some requirements occur across multiple phases, for example R7, “*Version control of files and other artefacts...*”.

### 4.3 Gap analysis

The gap analysis was guided by two key research questions (RQs).

- RQ1.** How can ModelOps be applied to resolve EMS requirements?
- RQ2.** Of the EMS requirements identified from the research, are there requirements that ModelOps is unlikely to address?

The assessment of an EMS requirement as Met or Unmet is based on an understanding of the current research gaps identified in the literature review and knowledge of ModelOps features. The aim is to ascertain whether ModelOps hypothetically addresses functional gaps and fulfils the requirements. The mapping below, between the 17 EMS requirements (from Sect. 4.2) and the 17 ModelOps features (from Table 1), addresses both RQs.

#### **Requirement R1. Discover, categorize and ingest increasing volumes of semi- and unstructured, multi-modal Big Data.**

- In theory, ModelOps feature F1 enables discoverability and federation of datasets and their relationships to other entities, using standard and emerging ontologies for IoT (smart devices, sensors, drones) and social media (Twitter, Facebook, etc.). This builds on increased use and consensus in vocabularies, ontologies and taxonomies (Hassani et al 2018; Kachaoui et al 2020; Thomas et al 2019; Asif et al 2021), improved IoT interoperability/standards (Arbib et al 2019; Moghaddam and Mucini 2019) and use of Open-source OWL (Kachaoui et al 2020). *In an emergency bushfire scenario, F1 enables discovery of relevant real-time Big Data from social media and multi-modal IoT sources using metadata and established taxonomies.*

- F2 enables automated ingestion of data from many sources and preparation of contextualised, “analysis-ready”, quality-coded data. This uses graph modeling (Beyer 2020) and indexed datasets to accelerate access and query operations (Barika et al 2019; Afyouni et al 2020). *In our scenario, once bushfire data are discovered, it must be ingested, contextualised and organised so that it is ready for use in the emergency situation at hand.*
- F3 facilitates maintaining high-quality and real-world datasets for training and testing. These can be datasets collated over long periods of time and from diverse scenarios (Buntain et al 2020; Glavic 2013). *A managed catalogue of relevant historical datasets for different categories, e.g., timestamped geo-spatial data of wildfires, climate conditions, dwellings, etc., can help categorize new data and train models.*
- F12 includes processes to manage dataset provenance, which enables historical queries and auditability of datasets (Hummer et al 2019), (Zhao et al 2015). *In the case of bushfire emergencies, F12 can play a crucial role in establishing trust in the provenance of multi-modal Big Data inputs.*

Combining F1 and F2 increases automation and efficiency of Big Data discovery, ingestion and pre-processing, which helps increase the speed of bushfire response, but it does not resolve the computation and storage resource issues that are outside the scope of R1. Overall, the gaps related to the implementation of R1 could be resolved by applying ModelOps features F1–F3 and F12, so the requirement is therefore assessed as “**Met**”.

**Requirement R2. Discover, categorize and ingest increasing volumes of satellite fire-event data.** Some of R2’s mapping is the same as R1, and the rationale for mapping to ModelOps features F1–F3 and F12 is also the same. In addition:

- F5 enables adoption of new algorithms / models for satellite data categorisation and analysis. This is important as satellite data volumes grow rapidly (Gozzard 2021) and researchers develop new bushfire models (Engel et al 2021; Goodrick 2021).
- F15 enables the comparison of existing “champion” algorithms / models against newly ingested “challenger” models. *It is important that new bushfire models are tested against existing models in a timely, robust manner.*

Overall, these features enable the categorization and ingest of an increasing volume of satellite data to the EMS. The gaps could be resolved by applying ModelOps features

F1–F3, F5, F12 and F15, so the requirement is therefore assessed as “Met”.

**Requirement R3. Create spatio-temporal and image datasets from unstructured social media data.** R3’s mapping is similar to that of R1. Specifically,

- F1 enables discoverability of social media posts that embed situational information (Thomas et al 2019; Daly and Thom 2016; Ahmady and Uchida 2020), including images posted by users (Alam et al 2018; Dunning and Breckon 2018). Geo-location of real-time images can increase the volume and geo-spatial coverage of sources of bushfire emergency data (Sun et al 2020; Buntain et al 2020).
- F2 enables spatio-temporal information from social media to be categorised using pre-processing ML algorithms (Pogrebnyakov and Maldonado 2017; Asif et al 2021; Dunning and Breckon 2018; Alam et al 2018) and continuously indexed (Afyouni et al 2020). *This feature increases the automation and efficiency of structuring social media data and images into indexed datasets and increases the volume of data available for fire-event model training and testing.*
- F12 enables historical queries and auditability of the new spatio-temporal disaster information datasets.

ModelOps features F1, F2 and F12 enable the categorization and ingest of previously “low-quality” social media data into auditable, re-useable datasets. R3 is assessed as “Met”.

**Requirement R4. Automate and rapidly classify information from phone / mobile calls.** R4 is a key requirement for managing the volume of calls from public and first responders to the emergency control-room.

- F3 enables ingestion of datasets for training and testing AI models, e.g. real-world voice data. *This is particularly useful for training on various accents and potentially in multi-lingual conditions.*
- F5 enables adoption of new algorithms / models for voice recognition. This feature enables Machine Listening algorithms to identify key words with a high priority and rapidly process real-time voice data using NLP (Ramchurn et al 2016).
- F15 enables the comparison of existing algorithms or champion models against newly ingested challenger models.

This requirement assumes improvements in linguistic models for voice recognition and less resistance from callers to emergency services responding to chatbot operators (Sun et al 2020). Overall, the gaps related to

the implementation of R4 could be resolved by applying ModelOps features F3, F5 and F15, so the requirement is therefore assessed as “Met”.

**Requirement R5: Effectively search very large spatio-temporal datasets and historical data.** This requirement is enabled by previous requirements for graph-enabled relationships between entities and dataset provenance. The intention is to enable a responder to quickly find a similar situation or previous incident, or ask AI for help to resolve a query.

- Feature F6 enables setup of decision tables or keyword tables for case-based reasoning (CBR). To increase efficiency, responders can use defined workflows or chatbots (Shahrah and Al-Mashari 2017).
- F7 determines who can manage the workflows or rules and make changes to them. Workflows or chatbots depend on user-intent mechanisms like keyword tables, decision trees or fuzzy search algorithms (Tsai et al 2019), that are auditable and secure.
- F8 enables the management of the lifecycle of decision-making and provenance of queries, commands or decisions. Theoretically, the system can learn and improve its performance whilst being auditable, increasing efficiency by presenting improvements to user-defined content (Shahrah and Al-Mashari 2017).

Overall, the gaps related to the implementation of R5 could be resolved by applying ModelOps features F6–F8, so the requirement is therefore assessed as “Met”.

**Requirement R6: Discover additional datasets where known volumes are insufficient to make valid predictions.**

- F1 enables discoverability and federation of datasets and their relationships to other entities, using standard and emerging ontologies for IoT, social media and remote sensing (satellites). The issue of insufficient data for machine-enabled decision-making is identified by Sun et.al (Sun et al 2020). By applying F1, AI could implement data discovery workflows.
- F6 enables AI to automatically apply business rules for identifying and categorising limitations on the veracity of predictions, e.g. “There is not enough data for a valid prediction”.

Overall, the gaps related to the implementation of R6 could be resolved by applying ModelOps features F1 and F6, so the requirement is therefore assessed as “Met”.

**Requirement R7: Version control of files and other artefacts to enable visualization of the history of changes.** R7



enables users to return to previous versions of assets used in EMS processes.

- F11 enables version control and documentation of datasets, models and model artefacts. This is important to enable the knowledge about a bushfire model to be packaged and transferred, making it easier to understand and re-use (Hummer et al 2019). The EC has raised requirements for traceability of critical AI decision-making (Commission European 2021).
- F12 includes processes to manage dataset provenance, which enables historical queries and auditability of datasets.
- F13 enables users to take a snapshot of a model in time, including all of the model's source code, artefacts, documentation, metadata and results. This enables dynamic responses, knowledge-sharing or testing of AI performance during emergency situations. Modellers can return to interim stages within modelling processes (Karamitsos et al 2020).

Overall, most gaps can be resolved. The requirement is **“Partially Met”**. None of the ModelOps features explicitly cover the visualisation of history of changes. F16 covers the visualisation of data, but not the history of changes.

**Requirement R8: Logging/tracking of decision-making activities. These may be interactive, iterative and involve new data, criteria or goals.** Some of R8's mapping is the same as R5, and the rationale for mapping to ModelOps features F6-F7 is also the same. In addition:

- F8 records the lifecycle of decision-making, including user queries, commands and decisions. Records are transparent and auditable, whether human or machine decisions.
- F9 enables an audit report that includes all steps in model execution and its results. An audit report enables assessment of whether bushfire EMS AI meets governance standards, e.g. the EC requirements for traceability, risk assessment and human oversight of machine-enabled decisions (Commission European 2021).

Overall, the gaps related to the implementation of R8 could be resolved by applying F6–F9, so the requirement is therefore assessed as **“Met”**.

**Requirement R9: Capture the processes, data, and parameters for experiments to become repeatable.** Some of R9's mapping is the same as R7, and the rationale for mapping to ModelOps features F11 and F13 is also the same. In addition:

- F10 enables viewing and management of models and all of their component parts, e.g. model code, coefficients, weights, etc. *It is beneficial for modellers to access a code repository like GitHub so code can be shared and assessed by a peer network.*
- F12 ensures that all model dependencies that are used to execute it, including data/tests/metrics, are managed and auditable. *It is important that fire-event modellers have access to quality training and test datasets.*
- F14 provides an interface to reproduce a model run, step by step. *In a dynamic bushfire environment, if the same model produces different results, users can quickly discover why.*

Overall, the gaps related to the implementation of R9 could be resolved by applying ModelOps features F10–F14, so the requirement is therefore assessed as **“Met”**.

**Requirement R10: Record and enable audit of complex, collaborative decision making.** R10's mapping is the same as R8, and the rationale for mapping to ModelOps features F7-F9 are also the same.

Li et al (2014) identified this gap and suggested methods to resolve it but recording and auditing of collaborative decisions in near real-time remains a challenge (Sun et al 2020). For example, record and audit complex decisions involving multiple approval levels, that are made outside the system or not using a system. Note that these might be handled in the EMS by enabling delegated approvals.

Overall, most gaps could be resolved by applying ModelOps features F7–F9 but the requirement is **“Partially Met”**.

**Requirement R11: Results from AI models should be explainable, repeatable, replicable.** This important requirement builds on requirements R7 and R9 and the rationale for mapping to ModelOps features F11, F12 and F14 is similar. In addition:

- F9 enables an audit report. *AI decisions can be interrogated at a later stage by expert user groups.*
- F11 enables documentation, so knowledge about a model can be packaged. *AI models can be packaged and shared amongst expert user groups.*
- F14 provides an interface to reproduce a model run, step by step. In bushfire event modelling, knowledge needs to be acquired and tested rapidly. Measures to improve the interpretability and explainability of AI models, such as explainable artificial intelligence (Arrieta et al 2020; Gunning et al 2019), could be built into stages in the modelling process, so that users can quickly compare results. Access to these artefacts enables expert users

to re-test AI models, after an event, for their performance (Commission European 2021).

Overall, the gaps related to the implementation of R11 could be resolved by applying ModelOps features F9, F11, F12 and F14, so the requirement is therefore assessed as “Met”.

**Requirement R12: Analyze social media to track feelings and reactions of the public; geo-spatial sentiment mining using natural language processing.** Although some researchers have concerns with the usefulness or quality of social media data (Basak et al 2020; McCreadie et al 2020; Thomas et al 2019), this requirement is in response to research that indicates social media can be used to call for help (Li et al 2019), identify human activity abnormalities and behaviours near disaster events (Zou et al 2019; Liu et al 2019) and assess psychological and healthcare needs (Kuang and Davison 2017). Post-event, Twitter data could be used to identify socio-geolocational disparities in the response effort (Zou et al 2019).

- F3 enables ingest of datasets for training and testing AI, using real-world social media data. *Real-time ingest of social media data enables real-time response to calls for help or recognising abnormal human activity. Data can be de-identified for training and testing.*
- F5 enables faster adoption of new algorithms / models (Li et al 2019; Zou et al 2019; Liu et al 2019) for categorisation and analysis of social media data.

The gaps related to the implementation of R12 could be resolved by applying ModelOps features F3 and F5, so the requirement is therefore assessed as “Met”.

**Requirement R13: Detect and manage false alarms.**

- F3 enables ingest of datasets for training and testing AI, using real-world data. *Training / testing data could be changed to produce “false positives” across a range of parameters.*
- F6 enables management of performance, including thresholds for statistical performance / accuracy and error messages. In a bushfire EMS, system administrators can set the triggers for alerts, acceptable levels of error and notification messages to control-room operators. An emerging approach to the detection of false alarms is using Uncertainty of Information (UoI) (Raglin et al 2021) to generate an uncertainty value for AI outputs.
- F8 records user queries and commands (decisions). *This includes the operator’s responses to false alarms. The operator response data could be re-used for training and testing.*

- F14 provides an interface to reproduce a model run, step by step. *If a model produces a false alarm, users can discover why.*

The gaps related to the implementation of R13 could be resolved by applying features F3, F6, F8 and F14, so the requirement is assessed as “Met”.

**Requirement R14: Re-configure and test existing workflows or simulations during an emergency event and re-deploy a solution.** This requirement addresses a gap identified with EMS systems (Hofmann et al 2015; Wagenknecht and Rueppel 2013) that has an inherent level of risk associated with deploying improvements to algorithms / models during an event.

- F3 enables ingest of new datasets for training and testing AI.
- F6 enables definition and management of rules, decision tables or keyword tables for compliance and performance. *In a bushfire scenario, all decisions and responses can be tracked.*
- F15 enables the comparison of existing champion workflows, algorithms or models against newly created challengers. *Improvements to workflows, algorithms or models can be tested in near real-time.*
- Feature F17 enables DevOps practices and tools for deploying model iterations, rollout and rollback. *Improvements that are approved by the control-room manager can be implemented or rolled back in near real-time.*

Although changes to the system during an emergency are high risk, the gaps could be resolved and the requirement is “Met”.

**Requirement R15: Test new fire-event algorithms or models as they are published.**

- F3 enables ingest of new datasets for training and testing AI.
- F4 enables a Standard Model definition with Active Metadata. Bushfire model discovery workflows might be enabled using Active Metadata (Beyer 2020).
- F5 enables adoption of new algorithms / models for data categorisation and analysis, which is important as researchers develop new models. EMS managers could setup a discovery process for new models and schedule it. This could enable, for example, discovery of bushfire simulation models that use coupled fire-atmosphere modelling (Goodrick 2021) or adaptations to the BRIGHT Machine Learning algorithm for re analysis (Biogeo-

graphical Region and Individual Geostationary Threshold) (Engel et al 2021).

Overall, the gaps related to the implementation of R15 could be resolved by applying features F3–F5, so the requirement is therefore assessed as “**Met**”.

**Requirement R16: User-friendly AI interfaces that require minimal technical expertise for practical use.** This requirement is important from a Human-Computer Interaction design perspective (Nielsen 2020). Several features mitigate the need for technical skills:

- F1 enables automated dataset discovery, reducing the need for direct dataset management.
- F6 enables an interface for defining and managing rules, decision tables or keyword tables for compliance and performance, setting thresholds for statistical performance/accuracy, setting custom/business KPIs, error messages and triggers for alerts and notification management.
- F14 enables an interface for step through of a model run, so that users can query why they have received these results.
- F16 provides a user console for model management and visualisations, with options for templates or custom reports.
- F17 enables tools for deploying model iterations, rollout and rollback. *It is likely that most EMS teams would need the support of a Data or Machine Learning Engineer for this function.*

Overall, the gaps related to the implementation of R15 could be partially mitigated by features F1, F6, F14, F16 and F17, so the requirement is therefore assessed as “**Partially Met**”.

**Requirement R17: Real-time support for tools that require a high level of competence in deployment.** An EMS is a mission-critical system that must provide information and enable decision-making in near real-time Sun et al (2020). Similar to R16, this requirement is important from a HCI design perspective (Nielsen 2020; De Silva 2018).

- Feature F6 enables defining and managing rules, decision tables or keyword tables, e.g. case-based reasoning (CBR) chatbots/workflows, which could assist with Q and A about system features. In future, bushfire EMS operators might be supported using intelligent agents (Damacharla et al 2020) or workflow bots.

This gap is not addressed by off-the-shelf ModelOps features and the requirement is therefore assessed as “**Unmet**”. In future, it might be addressed by new methods from Human-Machine Teaming (McNeese et al 2018) or Voice-Based

Synthetic Assistants (VBSA) (Damacharla et al 2020), or customising features such as F4–F6 for real-time operations in emergency control rooms.

In total, 13 of 17 requirements are assessed as Met, i.e. a Met ratio of > 75%. Three of the 17 requirements are partially met, and only one is unmet (R17). Based on these promising initial findings, further research into the application of ModelOps to EMS incident control-room operations is justified.

## 5 Evaluation of the EMS-ModelOps framework

### 5.1 Human-computer interaction assessment

Design reviews can be conducted on unfinished work or a set of specifications (Nielsen 2020). This may be a standalone design critique, where a focus group discuss the specification to determine whether it will meet its objectives. Alternatively, individuals can be asked to respond to a confidential survey, focusing on usability heuristics (Nielsen 2020).

**Focus groups:** Two or more focus groups of 5–6 participants could be recruited from EMS control-room operational teams to review and discuss the application of an EMS-ModelOps framework to bushfire EMS. Table 2 proposes three questions (Q1, Q2, Q3) to scope the group’s conversation, based on Flentge et al.’s (Flentge et al 2008) work with firefighters, incident managers and operators of critical infrastructures.

The format would be semi-unstructured, using open-ended questions. The advantage of a focus group is in-depth information, however the facilitator must be able to respond to participant’s queries about the EMS-ModelOps framework and how it might work as a prototype (Nielsen 2020).

**Confidential survey:** a survey of 20 or more participants could be recruited as per above. Participants would be provided with the research background and proposed framework. Using three open-ended qualitative questions (as per Table 2) and six closed quantitative questions with a 5-point Likert scale (very poor, poor, neither poor nor good, good, very good) to rate feature usability (Nielsen 2020), the survey would provide data to assess the participants expectation of Met or Unmet need.

In addition to design reviews, a longitudinal Ethnographic observation study could be used, following Kox and Lüder (2021). Their methods are relevant for bushfire EMS systems, where sophisticated emergency teams interact regards complex tasks and must coordinate and exchange real-time critical information. The study of incident control-room practices should be in a real-world setting, using a mixture of direct observation and interviews (Kox and Lüder 2021)

**Table 2** HCI survey questions adapted from Flentge et al. (Flentge et al 2008) and Nielsen (Nielsen 2020)

| Question  | Response type   |
|---|---|
| Q1. Will the proposed system provide information about the current situation and the availability of resources, by processing and filtering incoming information and by contributing to the assessment of the current situation (Flentge et al 2008)? | Open-ended text   |
| Q2. Will the proposed system provide advanced modelling, simulation, planning and analysis tools to improve the basis for decisions (Flentge et al 2008)?   | Open-ended text   |
| Q3. Will the proposed system provide capabilities to supply all involved parties with relevant information and improve command and control (Flentge et al 2008)?  | Open-ended text   |
| Q4. How do you rate the ability to keep users informed about what is occurring, through appropriate, timely information and feedback (Nielsen 2020)?  | (very poor, poor, neither poor nor good, good, very good) |
| Q5. If an action is performed by mistake or an error made, how do you rate the ability to diagnose the error and provide methods for its correction (Nielsen 2020)?   | (very poor, poor, neither poor nor good, good, very good) |
| Q6. How do you rate the ability to prevent errors (Nielsen 2020)?   | (very poor, poor, neither poor nor good, good, very good) |
| Q7. How do you rate the ability to precisely indicate the problem, and constructively suggest a solution (Nielsen 2020)?  | (very poor, poor, neither poor nor good, good, very good) |
| Q8. How do you rate the ability to make elements, actions, and options visible, reducing the need for memory (Nielsen 2020)?  | (very poor, poor, neither poor nor good, good, very good) |
| Q9. How do you rate the ability to provide information which is relevant, minimising information that is not useful (Nielsen 2020)?   | (very poor, poor, neither poor nor good, good, very good) |

to assess whether participants needs would be Met or Unmet under the proposed EMS-ModelOps framework.

## 5.2 Uncertainty of information hypotheses testing: prototype only

Assuming an EMS-ModelOps framework could be prototyped, testing could occur against a series of performance-related hypotheses. For example, does the prototype reduce the rate of errors? Are there less false alarms? Does it improve predictive capability?

Testers can apply the Uncertainty of Information (Raglin et al 2021) methodology. In the example below, testing uses concepts from Kothari (2004).

**H0:** Null Hypothesis - The uncertainty (error) in the EMS-ModelOps prototype is identical to the prototype without ModelOps technology.

**HA:** Alternative Hypothesis - The uncertainty (error) in the EMS-ModelOps prototype is less than the prototype without ModelOps technology.

**Test:** Run a selected UoI algorithm (Raglin et al 2021) in both prototypes to rank errors in the model output of “Model XYZ”. The independent variable is which prototype is in use: EMS-ModelOps or EMS without ModelOps. The dependent variable is the UoI algorithm result.

**Analysis:** Compare the rankings generated by the UoI computation.

**Interpretation:** If, the uncertainty (error) in the EMS-ModelOps prototype is less than the EMS prototype without ModelOps, then we have reduced the rate of

errors. Note that it is possible that the uncertainty (error) in the EMS-ModelOps prototype is more than the EMS prototype without ModelOps, if so, then we have increased the rate of errors.

## 6 Threats to validity

The review of bushfire EMS research and identification of issues have informed the development of the EMS-ModelOps framework and have been cited where applicable. The 17 EMS requirements are relatively high-level, as are the 17 ModelOps features that are mapped to them. It is possible that both lists are missing important elements, which are likely to be identified by other researchers.

Commercial experience and sector knowledge helped to identify which of the ModelOps features are applicable to the EMS requirements and whether they result in Met or Unmet, which introduces the possibility of personal bias.

The project cited published research from Gartner and online articles from Forbes about the application of ModelOps software in the private sector. There are no academic reviews of the quality of Gartner’s Information Technology research or their predictive capability.

There is a need to complete a wide-ranging review of XOps technologies, especially to assess what DataOps, ModelOps or DevOps features have been combined as XOps platform offerings. Access to commercial specifications would be useful.

To address potential gaps or criticism of the work, next steps could include (i) a survey of enterprise-level XOps systems, (e.g. Microsoft Azure, Databricks, MLflow, Kubeflow,

DataRobot, ModelOp.com, SAS, etc.) to assess their functionality; (ii) a high-level assessment of proprietary EMS applications against the EMS-ModelOps framework; or (iii) evaluation of the EMS-ModelOps framework by focus group discussion (Flentge et al 2008), survey of an expert user-group (Nielsen 2020) or observational study (Kox and Lüder 2021).

Given the framework is conceptual and experiments or tests were not conducted, there are no threats to the validity from issues with statistical methods, sampling, measurements or observations (Matthay and Glymour 2020).

## 7 Conclusion

Based on the research, the problem of managing huge, increasing volumes of heterogeneous data and complex “black box” AI/ML might be improved using ModelOps technology. The term black box is used to describe the problem of lack of human oversight of AI/ML decision-making and the European Commission has proposed legislation that critical software such as disaster management systems must enable such oversight. ModelOps, a nascent area of IT research, is increasingly common in the commercial sector and may be fit for this purpose.

The review of commercial sector research and investigation of ModelOps products enabled development of a shortlist of ModelOps features that theoretically enable automation of data management pipelines, model management and AI/ML governance. Based on the earlier evaluation of bushfire EMS, the novel EMS-ModelOps framework, if implemented, could theoretically resolve many of the identified issues. Mapping the EMS requirements to ModelOps features resulted in a greater than 75% Met ratio. Based on such promising initial findings, further research into the application of ModelOps to EMS is justified.

The analysis indicates that the EMS-ModelOps framework is generic and portable; it could be extended to other mission-critical applications through “bolted-on” technologies. To facilitate this, next steps should include (i) a survey of enterprise-level XOps systems to assess their capacity to integrate with existing proprietary applications and (ii) development of the EMS-ModelOps framework to a detailed technical level, with acceptance criteria.

During the literature review the project undertook a rapid review of the 2020 Office of the Chief Scientist report into Australia’s research and technology capabilities relevant to bushfire response, resilience, and recovery (Finkel, Alan and others 2020). Next steps could include a review of the Defence Science and Technology (DST) decision-support capability, with reference to XOps. It is unclear whether the DST’s “artificial intelligence and data fusion” capability

or “quick integration of open source or unclassified data” includes DataOps, MLOps or ModelOps capabilities.

Finally, advances in architecture that support information and communications infrastructure was out of scope but if successful, satellite-laser data transmission technology (Gozzard 2021) will increase the frequency and volume of received satellite images. New ML/AI algorithms will be developed to process these data. Goodrick (2021) has also reviewed new wildfire simulation models that use coupled fire-atmosphere modelling to improve real-time forecasting. To take advantage of improvements in these technologies and improve EMS situational awareness, it will be important to prioritise EMS-ModelOps requirement R15, “Test new fire-event algorithms or models as they are published”.

**Funding** Open Access funding enabled and organized by CAUL and its Member Institutions

## Declarations

**Conflict of Interest** The authors declare that they have no conflict of interest.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Ahmady SE, Uchida O (2020) Telegram-based chatbot application for foreign people in Japan to share disaster-related information in real-time. In: 5th International conference on computer and communication systems (ICCCS). IEEE, pp 177–181
- Afyouni I, Khan AS, Aghbari ZA (2020) Spatio-temporal event discovery in the big social data era. In: Proceedings of the 24th Symposium on International Database Engineering & Applications. ACM, pp 1–6
- Alam F, Ofli F, Imran M (2018) CrisisMMD: Multimodal twitter datasets from natural disasters. In: Proceedings of the international AAAI conference on web and social media
- Arbib C, Moghaddam MT, Muccini H (2018) (2019) IoT flows: a network flow model application to building evacuation. A view of operations research applications in Italy. Springer, pp 115–131
- Arrieta AB, Díaz-Rodríguez N, Del-Ser J, et al (2020) Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. Information Fusion 58:82–115. <https://www.sciencedirect.com/science/article/pii/S1566253519308103>

- Asif A, Khatoon S, Hasan MM et al (2021) Automatic analysis of social media images to identify disaster type and infer appropriate emergency response. *J Big Data* 8(1):1–28
- Bailey S, McCormick S, Bernstein E, et al (2021) State of ModelOps 2021. <https://business-of-data.com/reports/state-of-modelops-2021/>. Accessed 14 Aug, 2021
- Baltas C (2021) Greece PM blames devastating wildfires on climate change as ancient Olympia comes under threat. <https://www.abc.net.au/news/2021-08-06/greece-wildfires-blamed-on-climate-change-by-prime-minister/100354910>. Accessed 7 Aug, 2021
- Barika M, Garg S, Zomaya AY et al (2019) Orchestrating big data analysis workflows in the cloud: research challenges, survey, and future directions. *ACM Comput Surv* 52(5):95. <https://doi.org/10.1145/3332301>
- Basak J, Bhaumik P, Roy S, et al (2020) A Crowdsourcing based information system framework for coordinated disaster management and building community resilience. In: Proceedings of the 21st international conference on distributed computing and networking. ACM, p Article 33
- Berger N (2021a) California fires advance as heatwave blankets western US. <https://www.abc.net.au/news/2021-07-11/california-fires-advance-as-heatwave-blankets-western-us/100284324>. Accessed 18 Aug, 2021
- Berger N (2021b) Canadian and US officials brace for wildfires as heatwave continues. <https://www.abc.net.au/news/2021-07-01/canada-us-northwest-brace-for-wildfires-heatwave/100258250>. Accessed 18 Aug, 2021
- Berger N (2021c) US firefighters bracing for dangerous weather as 70 wildfires burn across California and Oregon. <https://www.abc.net.au/news/2021-07-18/us-wildfires-burn-across-california-oregon-dangerous-weather/100302756>. Accessed 18 Aug, 2021
- Beyer M (2020) Maximize Benefits in Your Improved Data Management - 2021 and Beyond. Gartner Webinar, <https://www.gartner.com/en/webinars/3988158/maximize-benefits-in-your-improved-data-management-2021-and-beyond>. Accessed 17 Aug, 2021
- Brethenoux E, Subramanyam J, Choudhary F, et al (2018) Gartner panel discussion: use ModelOps for an effective AI strategy. <https://www.gartner.com/en/webinars/3992789/panel-discussion-how-to-propel-ai-from-prototypes-to-scalable-va>. Accessed 14 Oct, 2020
- Buntain C, McCreddie R, Soboroff I (2020) Incident streams 2020: TRECIS in the time of COVID-19. Proceedings of the 28th text retrieval conference, Gaithersburg, Maryland, US
- Choudhary F, Vashisth S, Chandrasekaran A, et al (2020) Gartner's innovation insight for ModelOps. Accessed 20 Nov, 2020
- European Commission (2021) Europe fit for the digital age: Commission proposes new rules and actions for excellence and trust in artificial intelligence. Geneva, Switzerland, European Commission
- Daly S, Thom JA (2016) Mining and classifying image posts on social media to analyse fires. In: ISCRAM 13th international conference on information systems for crisis response and management, Citeseer, pp 1–14
- Damacharla P, Dhakal P, Bandreddi JP et al (2020) Novel human-in-the-loop (HIL) simulation method to study synthetic agents and standardize human-machine teams (HMT). *Appl Sci* (2076–3417) 10(23):8390
- DataRobot (2021) MLOps 101: the foundation for your AI strategy. <https://www.datarobot.com/ml-ops-101/>. Accessed 17 Aug, 2021
- De Silva N (2018) Role of human-computer interaction in disaster management. Investing in disaster risk reduction for resilience, annual symposium, pp 188–194
- Deacon B (2020) Zombie fires, fire thunderstorms hit Arctic as record heatwave heads into sixth month. <https://www.abc.net.au/news/2020-06-19/zombie-fires-and-fire-thunderstorms-hit-arctic/12369818>. Accessed 18 July, 2021
- Dugdale J, Moghaddam MT, Muccini H (2021) IoT4Emergency: internet of things for emergency management. Challenges and envisioned solutions. *SIGSOFT Softw Eng Notes* 46(1):33–36. <https://doi.org/10.1145/3437479.3437489>
- Dunnings AJ, Breckon TP (2018) Experimentally defined convolutional neural network architecture variants for non-temporal real-time fire detection. In: 2018 25th IEEE international conference on image processing (ICIP), IEEE, pp 1558–1562
- Engel CB, Jones SD, Reinke KJ (2021) Real-time detection of day-time and night-time fire hotspots from geostationary satellites. *Remote Sens* 13(9). <https://doi.org/10.3390/rs13091627>, <https://www.mdpi.com/2072-4292/13/9/1627>
- Feinberg D, Thanaraj R (2020) Gartner hype cycle for data management, 2020. Tech. Rep. G00450207, Gartner
- Field CB, Barros V, Stocker TF et al (2012) Managing the risks of extreme events and disasters to advance climate change adaptation: special report of the intergovernmental panel on climate change. Cambridge University Press, Cambridge
- Finkel Alan et al (2020) Bushfire research and technology: mapping Australia's capability. Tech. rep, Office of the Chief Scientist
- Flentge F, Weber SG, Behring A, et al (2008) Designing context-aware HCI for collaborative emergency management. In: Int'l Workshop on HCI for Emergencies in conjunction with CHI, vol 8. Citeseer
- Gartner (2021) Top 10 data and analytics technology trends for 2021. *Software World* 52(2):15. Issue 2, A.P. Publications Ltd
- Glavic B (2013) Big data provenance: challenges and implications for benchmarking. In: Specifying big data benchmarks: first workshop, WBDB 2012, San Jose, CA, USA, May 8–9, 2012 and second workshop, WBDB 2012, Pune, India, December 17–18, 2012, Revised Selected Papers, Springer, p 72
- Goodrick SL (2021) Special issue fire and the atmosphere. *Atmosphere* 12.1. <https://www.mdpi.com/2073-4433/12/1/66>
- Gozzard D (2021) Australia installs first space laser optical ground station in southern hemisphere. <https://www.abc.net.au/news/science/2021-09-12/australias-first-space-laser-communications-installed-wa/100406742>. Accessed 14 Sep, 2021
- Gunning D, Stefk M, Choi J, et al (2019) XAI-explainable artificial intelligence. *Sci Robot* 4:7120. <https://www.sciencedirect.com/science/article/pii/S1566253519308103>
- Hanly L, Sekar D (2021) Data science and machine learning at scale with Databricks (Webinar). Accessed 17 Sep, 2021
- Hassani A, Haghghi PD, Ling S, et al (2018) Querying IoT services: a smart carpark recommender use case. In: 2018 IEEE 4th world forum on internet of things (WF-IoT), IEEE, pp 619–624
- Hennessy K, Lucas C, Nicholls N et al (2005) Climate change impacts on fire-weather in south-east Australia. Climate impacts group, CSIRO Atmospheric Research and the Australian Government Bureau of Meteorology, Aspendale
- Hofmann M, Betke H, Sackmann S (2015) Process-oriented disaster response management: a structured literature review. *Bus Process Manag J* 21(5):966–987
- Hummer W, Muthusamy V, Rausch T, et al (2019) Modelops: Cloud-based lifecycle management for reliable and trusted ai. In: 2019 IEEE international conference on cloud engineering (IC2E), IEEE, pp 113–120
- Kachaoui J, Larioui J, Belangour A (2020) Towards an ontology proposal model in data lake for real-time COVID-19 cases prevention. *Int J Online Biomed Eng (iJOE)* 16(09):123
- Karamitsos I, Albarhami S, Apostolopoulos C (2020) Applying DevOps practices of continuous automation for machine learning. *Information* 11(7):363
- Kothari CR (2004) Research methodology: methods and techniques. New Age International, New Delhi

- Kox T, Lüder C (2021) Impacts as triggers for weather-related decision making: observations at the berlin fire brigade control and dispatch center. *Int J Disaster Risk Sci* 12(4):610–615
- Kua J, Loke SW, Arora C et al (2021) Internet of things in space: a review of opportunities and challenges from satellite-aided computing to digitally-enhanced space living. *Sensors* 21(23):8117
- Kuang S, Davison BD (2017) Learning word embeddings with Chi-Square weights for healthcare tweet classification. *Appl Sci* 7(8):846
- Kutukov R (2021) Wildfires rage across Yakutia, Siberia, threatening a power plant and disrupting transport. <https://www.abc.net.au/news/2021-07-20/siberia-huge-fires-yakutia/100305994>. Accessed 20 July, 2021
- Li J, Stephens KK, Zhu Y et al (2019) Using social media to call for help in Hurricane Harvey: bonding emotion, culture, and community relationships. *Int J Disaster Risk Reduct* 38(101):212
- Li N, Sun M, Bi Z et al (2014) A new methodology to support group decision-making for IoT-based emergency response systems. *Inf Syst Front* 16(5):953–977
- Li T, Xie N, Zeng C et al (2017) Data-driven techniques in disaster information management. *ACM Comput Surv* 50(1):1. <https://doi.org/10.1145/3017678>
- Liu Z, Du Y, Yi J et al (2019) Quantitative estimates of collective geo-tagged human activities in response to typhoon Hato using location-aware big data. *Int J Digit Earth* 13(4):1–21
- Matthay EC, Glymour MM (2020) A graphical catalog of threats to validity. *Epidemiology* 31.3:376–384. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7144753/>
- McCreadie R, Buntain C, Soboroff I (2020) Incident streams 2019: actionable insights and how to find them. CoRe paper—social media for disaster response and resilience. Proceedings of the 17th ISCRAM conference, Blacksburg, VA, USA
- McNeese NJ, Demir M, Cooke NJ et al (2018) Teaming with a synthetic teammate: insights into human-autonomy teaming. *Human Factors* 60(2):262–273
- ModelOp.com (2020) ModelOp.com RFP requirements template. <https://www.modelop.com/modelops-rfp/>. Accessed 13 Aug, 2021
- Moghaddam MT, Muccini H (2019) Fault-tolerant IoT. In: International workshop on software engineering for resilient systems, Springer, pp 67–84
- Moraffah R, Karami M, Guo R et al (2020) Causal interpretability for machine learning—problems methods and evaluation. *SIG-KDD Explor NewsL* 22(1):18–33. <https://doi.org/10.1145/3400051.3400058>
- Nielsen J (2020) 10 Usability heuristics for user interface design. <https://www.nngroup.com/articles/ten-usability-heuristics/>. Accessed 20 July, 2021
- Parry ML, Canziani O, Palutikof J et al (2007) Climate change 2007—impacts, adaptation and vulnerability: working group II contribution to the fourth assessment report of the IPCC, vol 4. Cambridge University Press
- Peacock S, Stevens R, Hollingworth K (2021) High-risk bushfire days in Victoria could triple by end of the century, CFA research finds. <https://www.abc.net.au/news/2021-08-16/extreme-fire-danger-days-could-triple-victoria/100380946>. Accessed 17 Aug, 2021
- Pogrebnyakov N, Maldonado E (2017). Identifying emergency stages in facebook posts of police departments with convolutional and recurrent neural networks and support vector machines. <https://doi.org/10.1109/BigData.2017.8258464>. IEEE international conference on big data, 2017
- Raglin AJ, Metu S, Lott D (2021) Decisions, graphs, and artificial reasoning for uncertainty of information. In: Pham T, Solomon L (eds) Artificial intelligence and machine learning for multi-domain operations applications III. International Society for Optics and Photonics, vol 11746. SPIE, pp 380–389, <https://doi.org/10.1117/12.2585517>
- Ramchurn SD, Huynh TD, Wu F et al (2016) A disaster response system based on human-agent collectives. *J Artif Intell Res* 57:661–708
- Routray SK, Tengshe R, Javali A, et al (2019) Satellite based IoT for mission critical applications. In: 2019 International conference on data science and communication (IconDSC). IEEE, pp 1–6
- Shahrah AY, Al-Mashari MA (2017) Emergency response systems: research directions and current challenges. In: Proceedings of the second international conference on internet of things, data and cloud computing. ACM, pp 1–6, Article 161, <https://doi.org/10.1145/3018896.3056778>
- Sicular S, Vashisth S (2020) Gartner hype cycle for artificial intelligence, 2020. Tech. Rep. G00448060, Gartner
- Sun W, Bocchini P, Davison BD (2020) Applications of artificial intelligence for disaster management. *Natural Hazards* 103.1(3):2631–2689. <https://doi.org/10.1007/s11069-020-04124-3>
- Teague B, Pascoe S, McLeod R (2010) The 2009 victorian bushfires royal commission final report: summary. Tech. rep., 2009 Victorian Bushfires Royal Commission
- Thomas C, McCreadie R, Ounis I (2019) Event tracker: a text analytics platform for use during disasters. In: Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval. ACM, p 1341–1344
- Tsai MH, Chen JY, Kang SC (2019) Ask Diana: a keyword-based chatbot system for water-related disaster management. *Water* 11(2):234. <https://www.mdpi.com/2073-4441/11/2/234>
- Vashisth S, Brethenoux E, Choudhary F, et al (2020a) Use Gartner’s 3-stage MLOps framework to successfully operationalize machine learning projects. Accessed 21 July, 2020
- Vashisth S, Linden A, Hare J, et al (2020b) Gartner hype cycle for data science and machine learning, 2020. Tech. Rep. G00450404, Gartner
- Wagenknecht A, Rueppel U (2013) Process driven emergency management platform for flood response integrating web GIS technologies. *Int J Civil Environ Eng* 13(5):1
- Wu J (2020a) Forbes cognitive world: ModelOps is the key to enterprise AI. <https://www.forbes.com/sites/cognitiveworld/2020/03/31/modelops-is-the-key-to-enterprise-ai/?sh=3f0b62606f5a>. Accessed 13, Sep, 2021
- Wu J (2020b) ModelOps is just the beginning of enterprise AI. <https://www.forbes.com/sites/junwu1/2020/10/09/modelops-is-just-the-beginning-of-enterprise-ai/?sh=77bf9b679ff3>. Accessed 13 Sep, 2021
- Zhao L (2021) Event prediction in the big data era: a systematic survey. *ACM Comput Surv* 54(5):94. <https://doi.org/10.1145/3450287>
- Zhao Y, Li Y, Raicu I et al (2015) Enabling scalable scientific workflow management in the Cloud. *Future Gener Comput Syst* 46:3–16
- Zhou Y, Yu Y, Ding B (2020) Towards MLOps: a case study of ML pipeline platform. In: 2020 International conference on artificial intelligence and computer engineering (ICAICE), pp 494–500, <https://doi.org/10.1109/ICAICE51518.2020.00102>
- Zou L, Lam NSN, Shams S et al (2019) Social and geographical disparities in Twitter use during Hurricane Harvey. *Int J Disaster Risk Reduct* 12:1300–1318
- Zumar C, Uhlenhuth K (2021) Announcing databricks autologging for automated ML experiment tracking. <https://databricks.com/blog/2021/08/27/announcing-databricks-autologging-for-automated-ml-experiment-tracking.html>. Accessed 28 Aug, 2021