Studying Logging Practice in Machine Learning-based Applications





GIGL, Polytechnique Montréal

By Patrick Loic Foalem, Foutse Khomh, Heng Li



OUTLINE

- INTRODUCTION
- RELATED WORK AND CONTRIBUTIONS
- STUDY DESIGN
- RESEARCH QUESTIONS
- RESULTS AND FINDINGS
- CONCLUSION



INTRODUCTION

What is logging in software engineering?

★ Common practice in software engineering is to record runtime system information.



Logging code



System in production



Log messages

```
logging.info("Training accuracy {:.2f}" .format(accuracy))
```

Verbosity Level

Static Message

Dynamic Variable

```
2022-02-01 16:10:08 INFO Training accuracy 0.79
```

Why logging is important?

★ log is often used for various tasks in software maintenance and operations.



Debugging



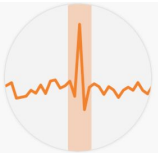
Log messages



Performance Analysis



Monitoring



Anomaly detection

Is logging within applications challenging?

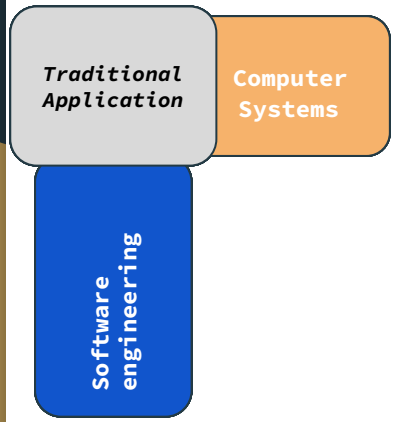


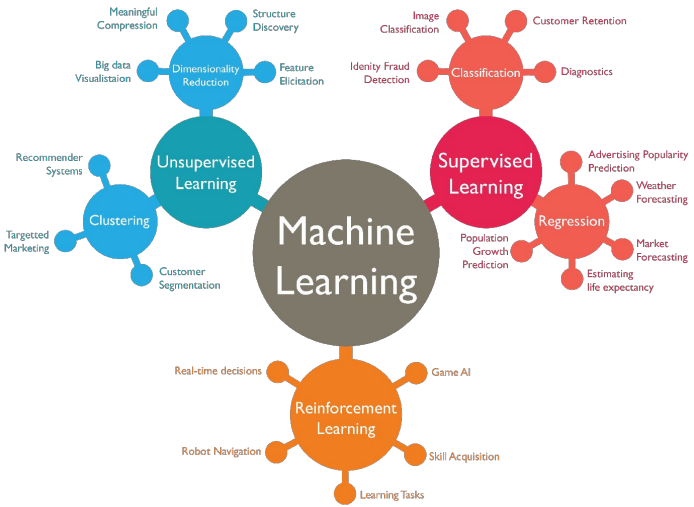


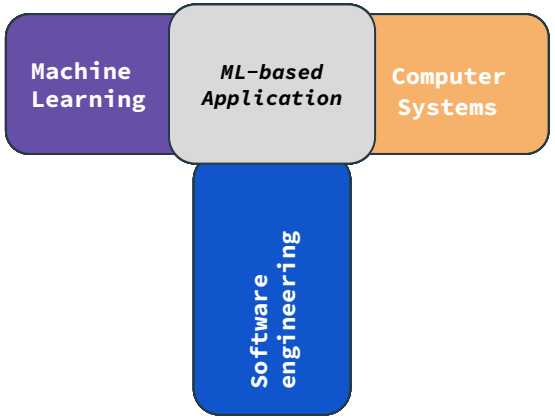
Too **much** : runtime overhead, storage overhead, too many trivial logs.

Too **little**: missing important information, increasing the difficulty for problem diagnose

From traditional application to ML-based application







Before

After

Challenge when developing ML-based application.

Engineering AI Systems: A Research Agenda

Jan Bosch
Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden
jan.bosch@chalmers.se

Helena Holmstrm Olsson
Computer Science and Media Technology
Malm University
Malm, Sweden
helena.holmstrom.olsson@mau.se

Ivica Crnkovic
Computer Science and Engineering
Chalmers University of Technology
Gothenburg, Sweden
ivica.crnkovic@chalmers.se

- Data drift
- Reproducibility of results
- Overfitting
- Monitoring and Logging

identified problems		strategic focus
Lack of labelled data	Data drift	data quality management
Lack of metadata	Data dependencies	
Shortage of diverse samples	Managing categorical data	
Heterogeneity in data	Managing sequences in data	
Data granularity	Deduplication complexity	
Imbalanced data sets	Data streams for training	
Experiment management	Lack of modularity	design methods and processes
Dependency management	Sharing and tracking techn.	
Unintended feedback loops	Reproducibility of results	
Effort estimation	Data extraction methods	
Cultural differences	Tooling	
Specifying desired outcome		
Overfitting	Limited transparency	model performance
Scalable ML pipeline	Training/serving skew	
Quality attributes	Sliced analysis of final model	
Statistical Understanding		
Monitoring and Logging	Privacy and data safety	deployment & compliance
Testing	Data silos	
Troubleshooting	Data storage	
Data sources and distribution	Resource limitations	
Glue code and support		

Research objectives



Understand logging practices in ML-based applications.



Compare ML logging with traditional logging.



Explore the motivation behind logging in ML-based projects.



RELATED WORK AND CONTRIBUTIONS

Characterizing logging practice in traditional application.

★ Logging practice is widely studied in traditional software.

Studying the characteristics of logging practices in mobile apps: a case study on F-Droid

Yi Zeng[1] · Jinfu Chen[1] · Weiyi Shang[1] · Tse-Hsun (Peter) Chen[1]

Characterizing logging practices in Java-based open source software projects – a replication study in Apache Software Foundation

Boyuan Chen[1] · Zhen Ming (Jack) Jiang[1]

Characterizing Logging Practices in Open-Source Software

Ding Yuan‡†, Soyeon Park†, and Yuanyuan Zhou†
†*University of California, San Diego,* ‡*University of Illinois at Urbana-Champaign*
{*diyuan,soyeon,yyzhou*}*@cs.ucsd.edu*

Where Do Developers Log?
An Empirical Study on Logging Practices in Industry

Qiang Fu[1], Jieming Zhu[2], Wenlu Hu[3], Jian-Guang Lou[1], Rui Ding[1], Qingwei Lin[1], Dongmei Zhang[1], Tao Xie[4]

Augmenting logging decisions using AI.

★ **Using AI algorithms to helps developers in their logging decisions.**

Learning to Log: Helping Developers Make Informed Logging Decisions

Jieming Zhu†, Pinjia He†, Qiang Fu§, Hongyu Zhang‡, Michael R. Lyu†, Dongmei Zhang‡
†Shenzhen Research Institute, The Chinese University of Hong Kong, Shenzhen, China
†Ministry of Education Key Laboratory of High Confidence Software Technologies (CUHK Sub-Lab)
†Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong
§Microsoft, Washington DC, USA ‡Microsoft Research, Beijing, China
†{jmzhu, pjhe, lyu}@cse.cuhk.edu.hk §qifu@microsoft.com ‡{honzhang, dongmeiz}@microsoft.com

Where Shall We Log? Studying and Suggesting Logging Locations in Code Blocks

Zhenhao Li
l_zhenha@encs.concordia.ca
Concordia University
Montreal, Quebec, Canada

Tse-Hsun (Peter) Chen
peterc@encs.concordia.ca
Concordia University
Montreal, Quebec, Canada

Weiyi Shang
shang@encs.concordia.ca
Concordia University
Montreal, Quebec, Canada

Which Log Level Should Developers Choose for a New Logging Statement?

Heng Li · Weiyi Shang · Ahmed E. Hassan

Using Deep Learning to Generate Complete Log Statements

Antonio Mastropaolo
antonio.mastropaolo@usi.ch
SEART @ Software Institute
Università della Svizzera italiana
Switzerland

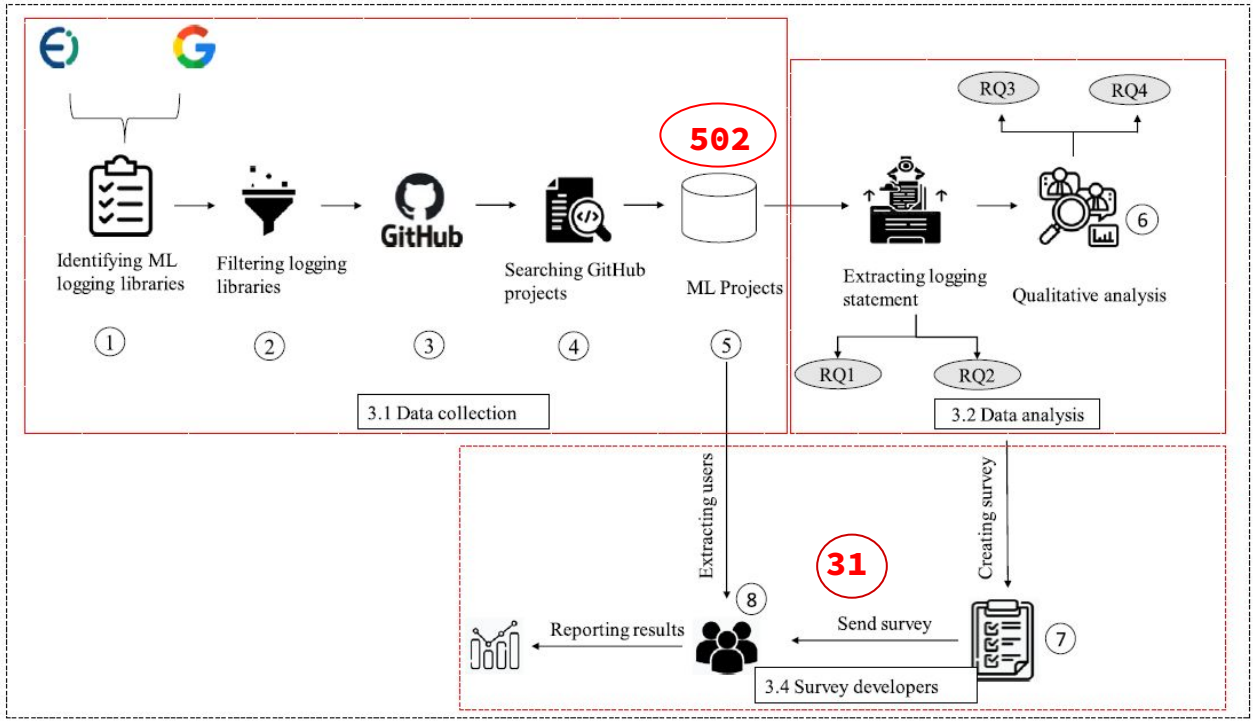
Luca Pascarella
luca.pascarella@usi.ch
SEART @ Software Institute
Università della Svizzera italiana
Switzerland

Gabriele Bavota
gabriele.bavota@usi.ch
SEART @ Software Institute
Università della Svizzera italiana
Switzerland



STUDY DESIGN

Empirical Study



Replication package
Try it out!



RESEARCH QUESTION

Four research questions.



★ **RQ1:** What are the characteristics of logging practices in ML-based software?



★ **RQ2:** What are the most commonly used ML-specific logging libraries in ML-based applications?



★ **RQ3:** Which phases of the ML pipeline are more prone to logging?



★ **RQ4:** Why do ML practitioners use logging in ML-based applications?



RESULTS AND FINDINGS

RQ1: What are the characteristics of logging practices in ML-based software?

Result (1): Prevalence of Logging in ML vs. Traditional Applications

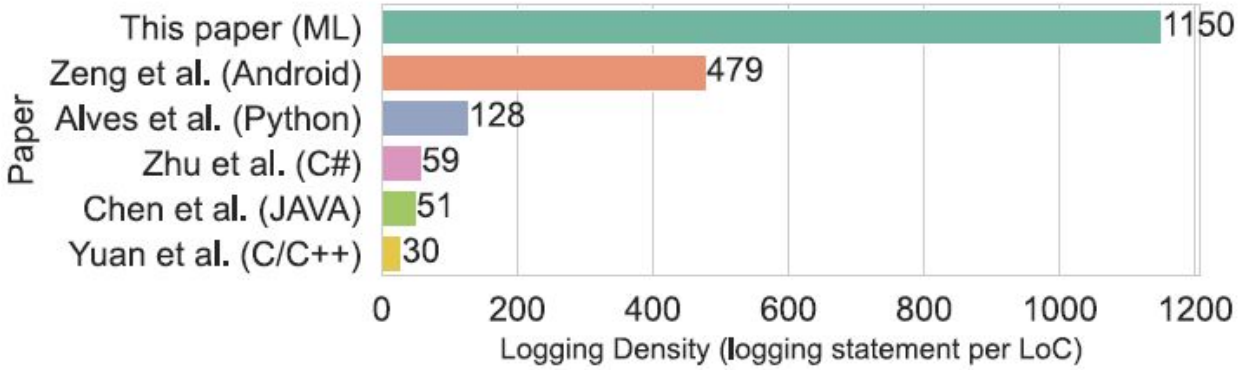


Fig. 4. Logging density across different papers.

Metric:
Log density = SLOC/NL

Finding 1: Our study reveals that logging practices in ML-based applications are less prevalent compared to traditional software systems.

RQ1: What are the characteristics of logging practices in ML-based software?

Result (2): Logging libraries types.

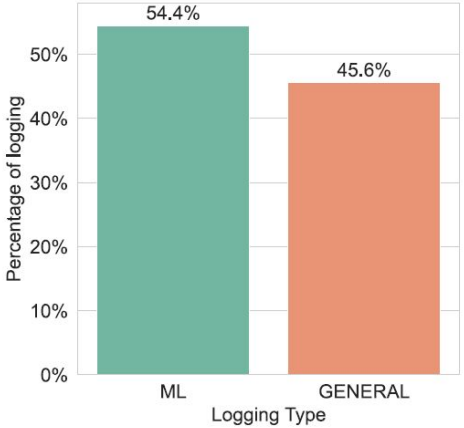


Fig. 6. Logging type distribution.

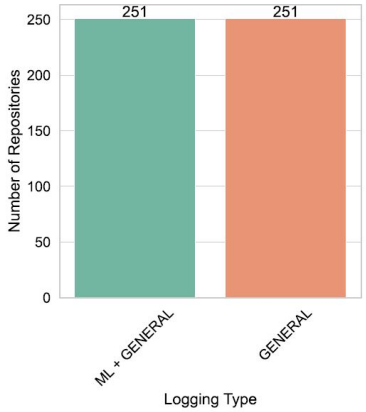


Fig. 7. Distribution of number of repository by logging type.

Logging libraries type	Libraries
ML-specific logging	Comet_ml Whylogs Wandb Tensorboard MLflow Tensorflow Neptune Dowel Sacred Ml-logger
General logging libraries	Logging Warnings

60% of 31 practitioners

Finding 2:

- Traditional applications primarily rely on general logging libraries.
- ML-based applications use a combination of ML-specific and general logging libraries.

RQ1: What are the characteristics of logging practices in ML-based software?

Result (3): Log levels in ML vs Traditional application

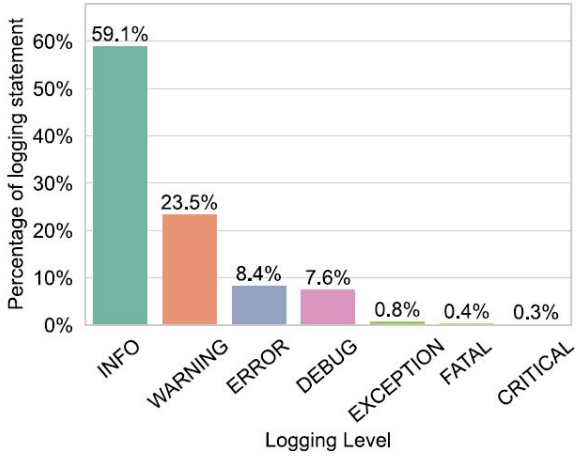


Fig. 8. Distribution of the logging statement levels.

- Logging model parameters, performance (INFO)

```
6       else:
7           logger.info('Epoch {} Loss {:.4f} Accuracy {:.4f}'.format
          (
8                 epoch, loss_meter.avg, accuracy))
```

- Data inconsistencies (Warning)

```
17  def dataset_from_dicts(self, dicts, indices=None, return_baskets=
        False, non_initial_token="X")
18                      logger.warning(f"[Task: {task_name}] Could
        not convert labels to ids via label_list!"
19                                     f"\nWe found a problem with
        labels {str(problematic_labels)}")
```

Finding 3: Info and warn log levels are the most commonly used in ML-based applications.

RQ2: What are the most commonly used ML-specific logging libraries in ML-based applications?

Result (1): Most Common ML-Specific Libraries

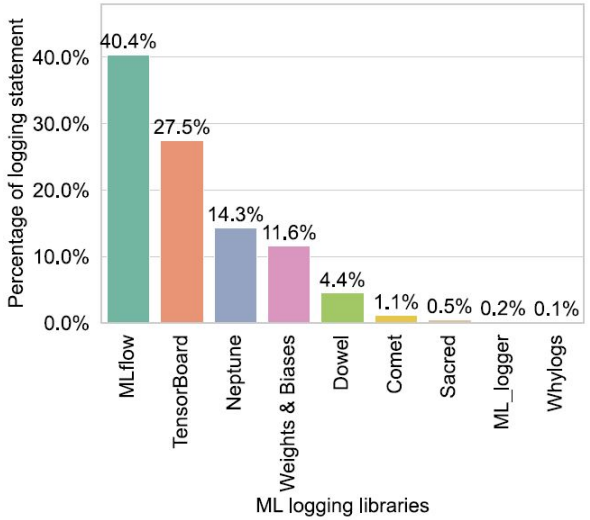


Fig. 10. Most commonly used ML-specific logging libraries.

- These libraries are designed to handle the unique requirements of ML logging, including performance metrics, model hyperparameters, and experiment tracking.

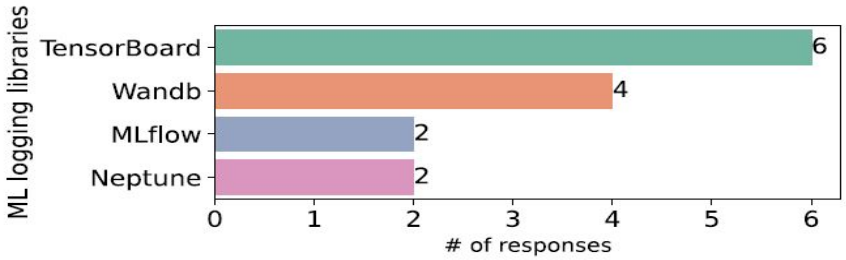


Fig. 9. Survey response on logging libraries used by practitioners.

Finding 3: The most commonly used ML-specific logging libraries are MLflow, TensorBoard, Neptune, and Weights & Biases.

RQ3: Which phases of the ML pipeline are more prone to logging?

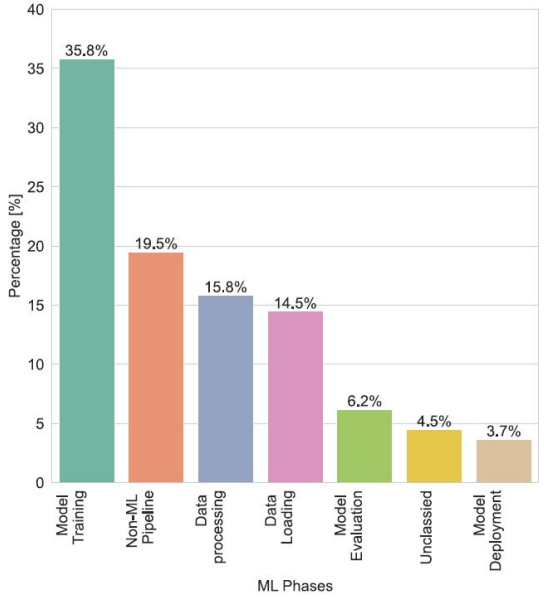


Fig. 12. Percentage of logging in different ML phases.

- Model training: need for extensive experimentation and configuration adjustments.
- Model Deployment: monitoring inference performance and detecting anomalies.

Finding 3: All five phases of the entire ML pipeline contain logging statements. However, the **model training** phase has the largest proportion of logging statements and the **model deployment** phase has the smallest proportion of logging statements.

RQ4: Why do ML practitioners use logging in ML-based applications?

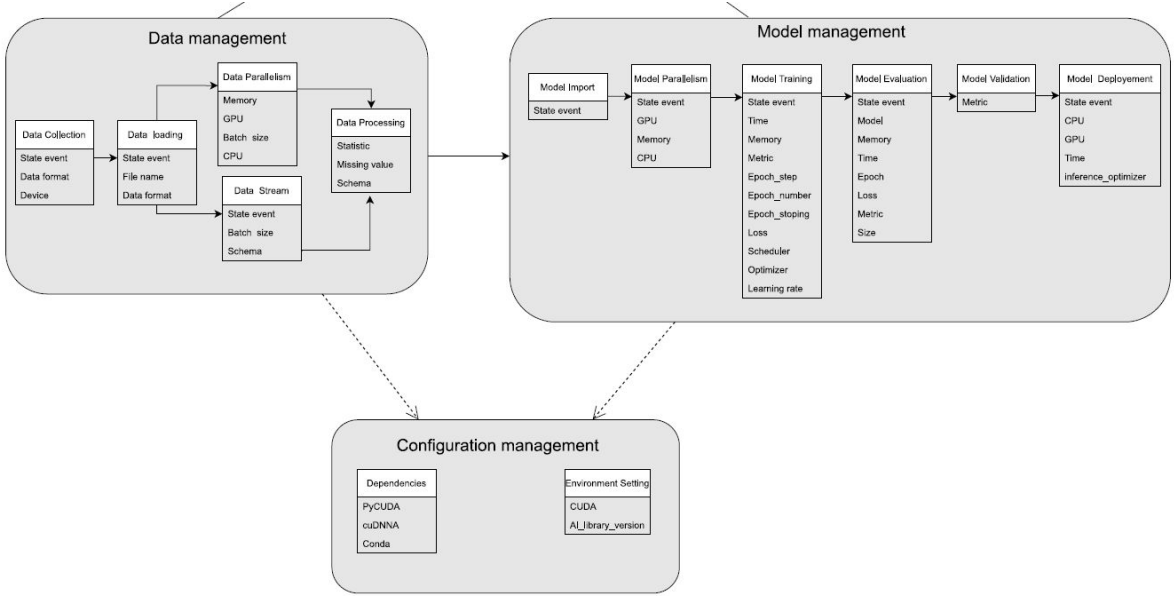


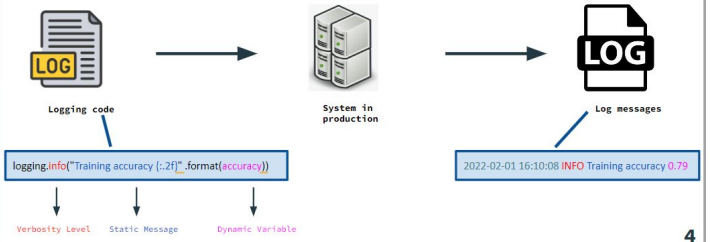
Fig. 13. Reason for Logging in ML-based application. *Entity represent different stage in ML-based system. Attributes represent what type or what information is logged.*

Finding 4: ML practitioners use logging statements to record important information in data management (e.g., recording the data format in data collection), model management (recording the hyperparameters in model training), and configuration management (e.g., recording the used resources or libraries).

CONCLUSIONS

What is logging in software engineering?

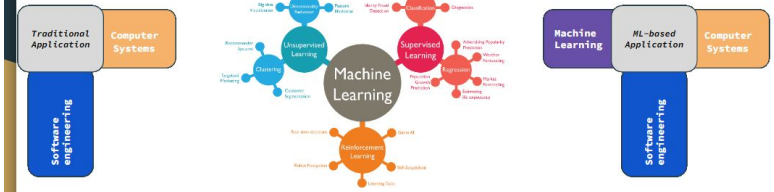
★ Common practice in software engineering is to record runtime system information.



4



From traditional application to ML-based application



Before

After

7

Four research questions.

★ RQ1: What are the characteristics of logging practices in ML-based software?

★ RQ2: What are the most commonly used ML-specific logging libraries in ML-based applications?

★ RQ3: Which phases of the ML pipeline are more prone to logging?

★ RQ4: Why do ML practitioners use logging in ML-based applications?

16

RQ1: What are the characteristics of logging practices in ML-based software?

Result (2): Logging libraries types.

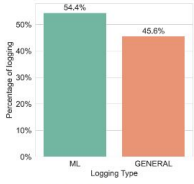


Fig. 6. Logging type distribution.

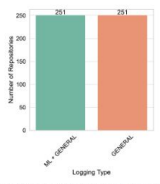


Fig. 7. Distribution of number of repository by logging type.

Logging libraries type	Libraries
ML-specific logging	Comet_ml Whylogs Wandb Tensorboard MLflow Tensorflow Neptune Dowel Sacred Ml-logger
General logging libraries	Logging Warnings

60% of 31 practitioners

Finding 2:

- Traditional applications primarily rely on general logging libraries.
- ML-based applications use a combination of ML-specific and general logging libraries.

19