

---

## A Supplementary material for the Execution Plan with the SE Authors

In this section of the Appendix, we provide the relevant information for the supplementary material used to survey the authors of the identified Software Engineering (Software Engineering (SE)) research papers using Machine Learning (Machine Learning (ML)). We sent emails to the authors asking them to answer the following question based on a subset of the identified papers in which they were authors, which was presented in the email.

**Question:** *What machine learning best practices have you used in your software engineering research papers, and how often did you apply those practices (Always, Frequent, Sometimes, Never)?*

**Instruction given bellow the question:**

*Please write each practice in a new line:*

- Good practice [how often]

## B Supplementary material for the Execution Plan with SE Researchers

In this section of the Appendix, we provide the relevant information for the supplementary material used to survey and interview the Software Engineering (SE) researchers that have used Machine Learning (ML).

### B.1 Demographic Survey Questions

We collected and report demographic data about our interview participants. This data is not reported within this article, but can be accessed in our online appendix. For simplicity, we present each question following the conventions below: multiple choice question with a single answer (○), multiple choice question with multiple answers (□), open question with a text field for answering (  ), questions with a numeric field with accepted answers between 0 and 100 (  # ). In addition, if clarifications or instructions were provided for a question, they will be presented after the question in italics between parentheses.

- i) *How many years of experience in software development do you have?*  #
- ii) *How many papers have you published related to ML? (ML: Machine Learning)*  #
- iii) *How many papers have you published related to SE without involving ML? (ML: Machine Learning, SE: Software Engineering)*  #
- iv) *How many papers have you published related to ML4SE? (ML: Machine Learning, SE: Software Engineering, ML4SE: ML for SE)*  #
- v) *How many papers have you published related to SE4ML? (ML: Machine Learning, SE: Software Engineering, SE4ML: SE for ML)*  #
- vi) *What venues (journals/conferences) do you tend/prefer to publish in? (Please, separate each venue with a comma ( , ))*
- vii) *How would you describe your primary role?*
  - ☐ Researcher (academics) ☐ Researcher (industrial)
  - ☐ Applied Scientist ☐ Tester ☐ Project Manager ☐ DevOps Engineer
  - ☐ Programmer ☐ Project Lead ☐ IT Manager ☐ Other
- viii) *What is your highest level of education?*
  - ☐ Did not graduate from high school ☐ Some college ☐ High school
  - ☐ Bachelors' degree ☐ Master's degree ☐ Doctoral degree ☐ Other

- 
- ix) Which programming languages have you frequently used in past project(s)? (If more than an additional language is added, please split them with “,”)
- ☐ C/C++      ☐ C#      ☐ Java      ☐ Javascript
- ☐ Julia      ☐ Python      ☐ R      ☐ Rust
- ☐ Dart      ☐ Kotlin      ☐ Other
- x) Which types of Learning approaches have you used?
- ☐ Supervised      ☐ Semi-supervised      ☐ Reinforcement learning
- ☐ Unsupervised      ☐ Other
- xi) What types of ML tasks have you used?
- ☐ Regression      ☐ Clustering      ☐ Machine translation
- ☐ Object detection      ☐ Transcription      ☐ Anomaly detection
- ☐ Other
- xii) What types of systems have you developed involving ML?
- ☐ Operating systems      ☐ Server side      ☐ Mobile applications
- ☐ Middleware      ☐ Desktop applications      ☐ Web applications
- ☐ Databases      ☐ Recommender systems
- ☐ Development tools (compilers, prog.languages, etc.)      ☐ Other
- xiii) Have you primarily worked on open-source or closed-source projects?
- ☐ Open-source only    ☐ Closed-source only    ☐ Open and closed-source
- xiv) For which domains (e.g., banking or healthcare) have you developed applications/systems? (List them next separated by comma ( , ))
- xv) Do you have a background in machine learning? (Formal (college classes, degree, certification) or informal (self-learning or other training)?)
- ☐ Yes, formal      ☐ Yes, informal      ☐ No
- xvi) Do you have a background in software engineering research? (Formal (college classes, degree, certification) or informal (self-learning or other training)?)
- ☐ Yes, formal      ☐ Yes, informal      ☐ No
- xvii) When researching software engineering, do you keep track of your approaches/ models + dataset?, Which versioning systems have you used (e.g., CML, DVC, Git)?
- ☐ Yes       ☐ No
- xviii) How often did you release or help release a new major version of ML models over the past two years? Please give your best estimate.
- ☐ Never      ☐ Annually      ☐ Quarterly      ☐ Monthly      ☐ More frequently
- xix) Which country/countries are you and/or your organization based in?

## B.2 Interview Questions

- i) Best Practices and Protocols
- What do you consider as a best practice?
  - Which process, guidelines, or pipelines you follow when employing ML in SE?
  - What do you consider are the “must” practices or protocols you use or implement in your research/papers?
  - Which are the most recurrent challenges when using ML for SE in any of the ML workflow stages?
  - Which practices do you use from each ML pipeline stage?
  - How did you employ the components of learning to guide your ML model design? [Abu-Mustafa, 2012]
  - Do you take into consideration any of the learning principles (e.g., data snooping, Occam’s razor, or sampling bias)? [Abu-Mustafa, 2012]
  - Which quality attributes are important for SE systems that use ML? *⚡* **ANSW1**

- 
- Do you have any particular difficulties/challenges when building an ML-enabled system to ensure **ANSW1** attributes?
  - ii) Education
    - What educational resources have you employed to perform ML4SE? (*e.g.*, tools, languages, books, etc..)
    - How do you learn ML4SE?
    - How do you educate your students to enable ML4SE?
    - How do you promote **ANSW1** when teaching about ML4SE?
  - iii) Reviewer's Perspective
    - What issues have you observed when reviewing papers?
    - What ML4SE areas are not being covered in conference reviews?
    - As a reviewer, have you seen that the **ANSW1** attributes are being addressed?
  - iv) ML Workflow Stages (*Amershi et al.*)
    - Model Requirements
      - What was the rationale of selecting or proposing an ML model you have utilized in your papers?
    - Data Collection
      - Do you collect your own data, use previous datasets, or both? How do you select data/datasets?
      - Which data collection strategies or protocols you employ in your studies?
      - How do you promote **ANSW1** when collecting data for ML-enabled systems?
    - Data Cleaning
      - Which pipelines do you employ to address data exploration?
      - How do you perform data cleaning?
      - How do you handle exploratory analysis in your studies?
      - How important are exploratory analyses for data cleaning?
      - How do you promote **ANSW1** when cleaning data in ML-enabled systems?
    - Data Labeling (if applicable)
      - Which data labeling protocols do you use for your supervised tasks?
      - How do you promote **ANSW1** when labeling data for ML-enabled systems?
    - Feature Engineering
      - Which feature engineering methods you have employed?
      - How do you promote **ANSW1** when selecting and building features in ML-enabled systems?
    - Model Training
      - What frameworks have you employed for model training?
      - How do you parameterize your models?
      - How do you promote **ANSW1** when training models in ML-enabled systems?
    - Model Evaluation
      - How do you evaluate or validate your ML models?
      - Do you employ any measurements to control for bias?
      - Do you use any interpretability technique to guide your evaluation?
      - How are measurements aligned to study RQs, goals, or business objectives?
      - How do you promote **ANSW1** when evaluating ML-enabled systems?
    - Model Deployment
      - How do you do promote **ANSW1** when deploying ML-enabled systems?
    - Model Monitoring
      - Do you employ any measurements to handle concept drift?
      - How do you monitor models during operation (*e.g.*, user studies, industry projects, production)?
      - How do you promote **ANSW1** when monitoring ML-enabled systems?
  - v) Potential Follow-up Questions

## C Supplementary Material for the Results of the Open Coding, 2011-2022

### C.1 Codes associated to the ML practices

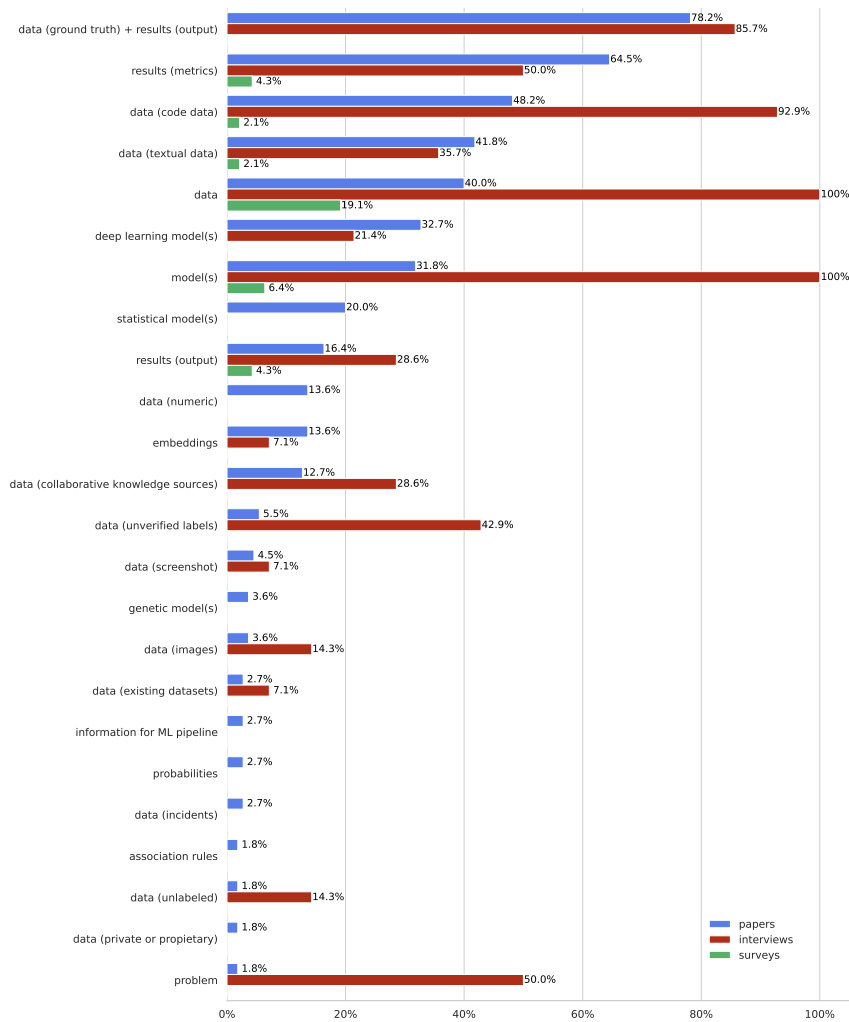


**Fig. 1** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *input category* appears. The figure displays the first half of the *input codes*, organized by the subject *papers* appearance.

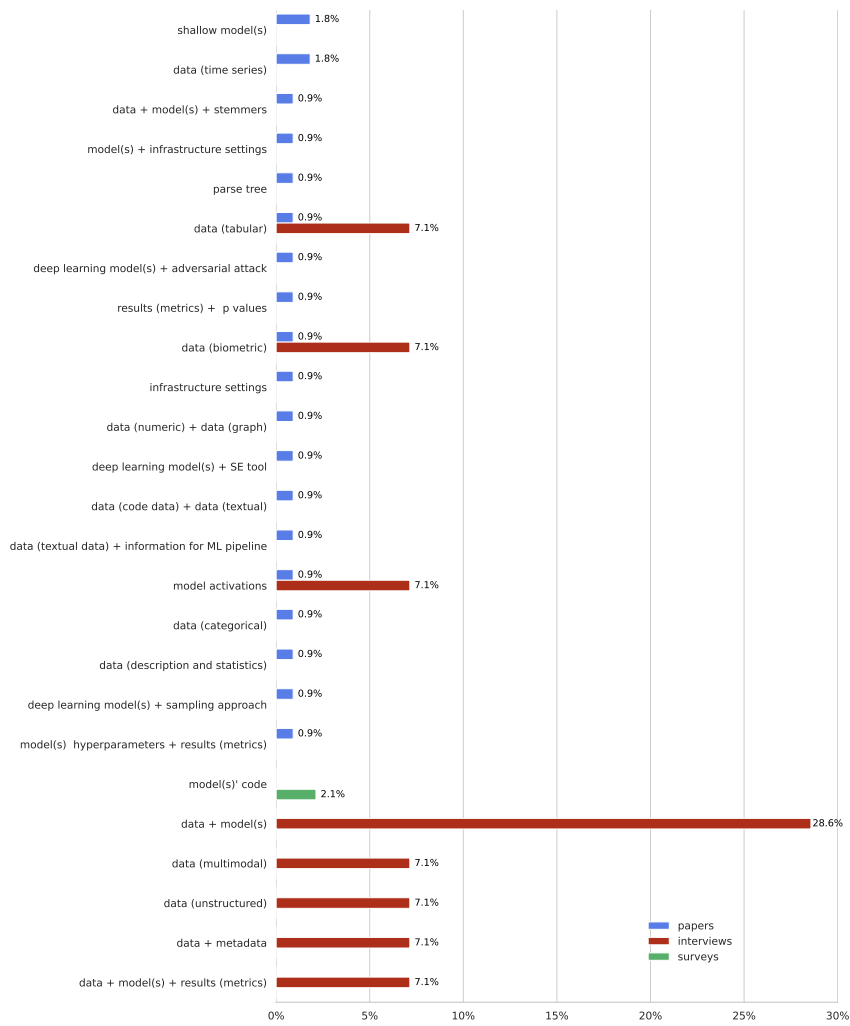


**Fig. 2** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *input category* appears. The figure displays the second half of the *input codes*, organized by the subject *papers* appearance.

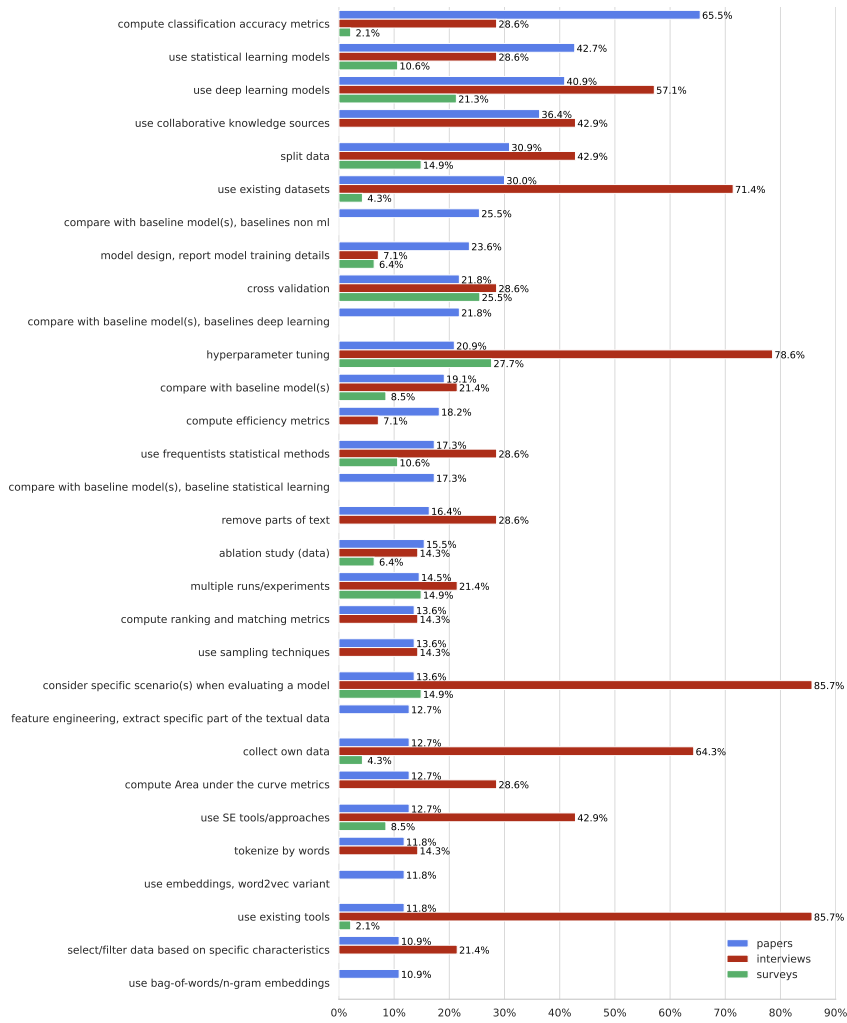


**Fig. 3** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the first sixth of the *technique codes*, organized by the subject *papers* appearance.

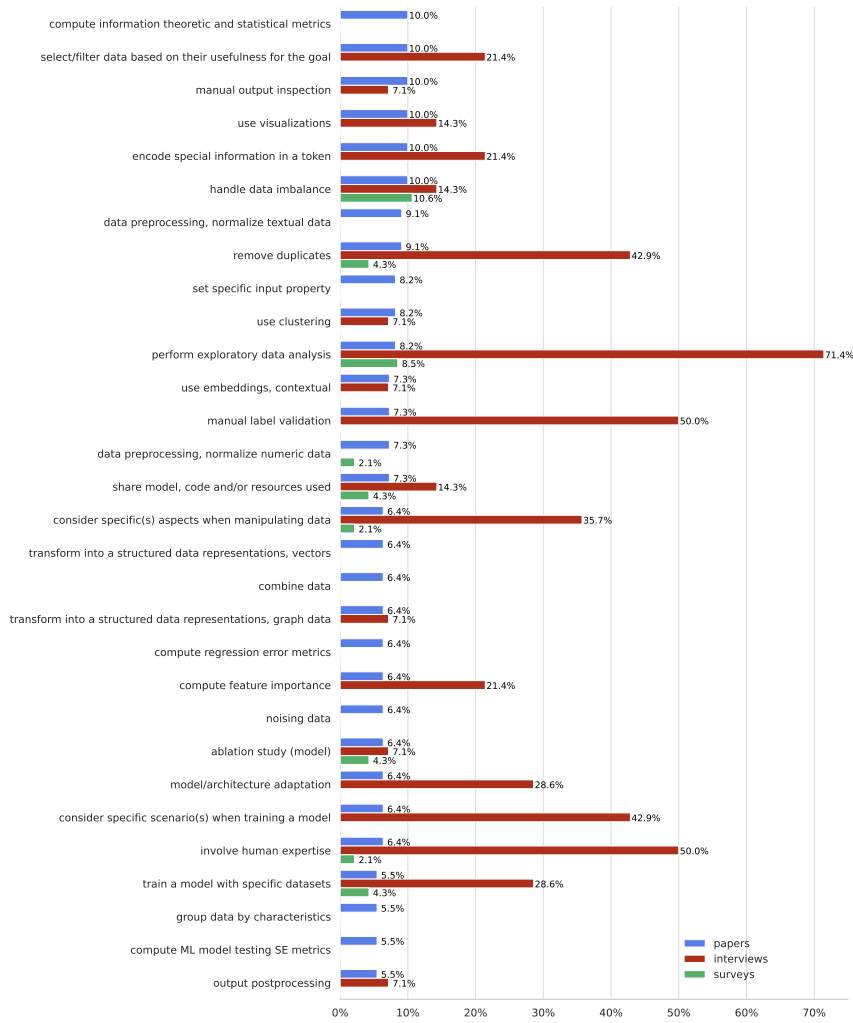


**Fig. 4** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the second sixth of the *technique codes*, organized by the subject *papers* appearance.

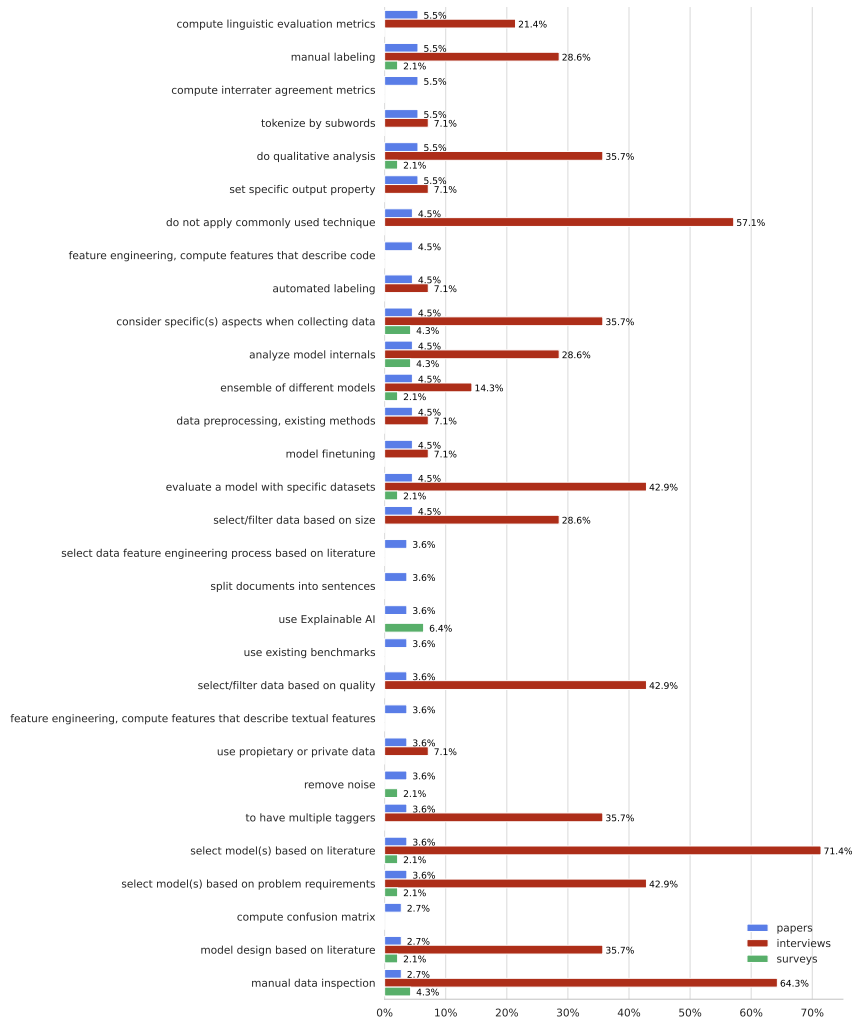


**Fig. 5** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the third sixth of the *technique codes*, organized by the subject *papers* appearance.

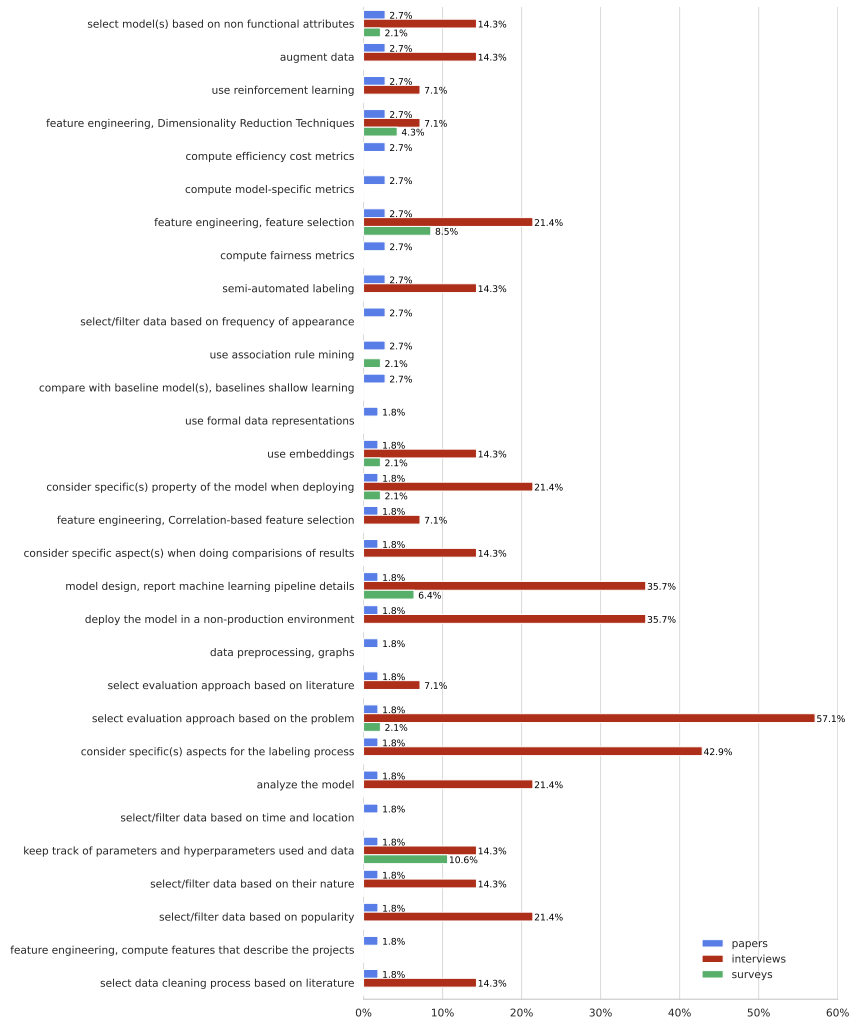


**Fig. 6** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the fourth sixth of the *technique codes*, organized by the subject *papers* appearance.

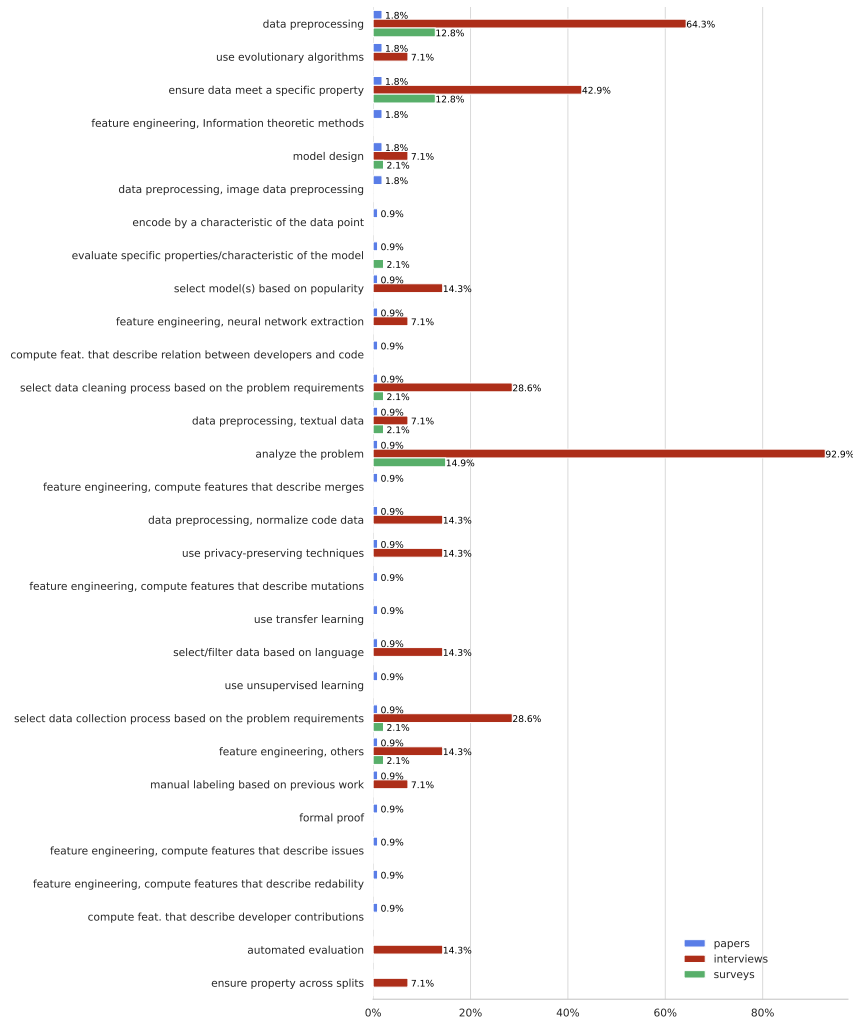


**Fig. 7** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the fifth sixth of the *technique codes*, organized by the subject *papers* appearance.

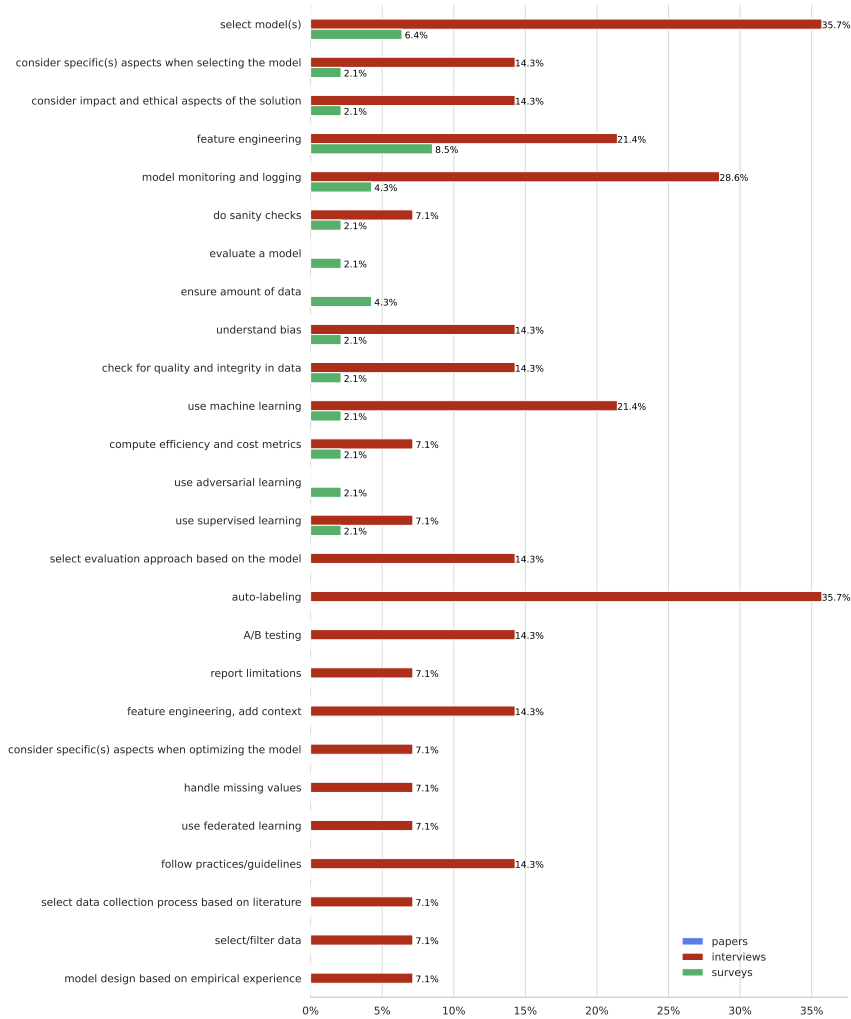


**Fig. 8** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *technique category* appears. The figure displays approximately the last sixth of the *technique codes*, organized by the subject *papers* appearance.

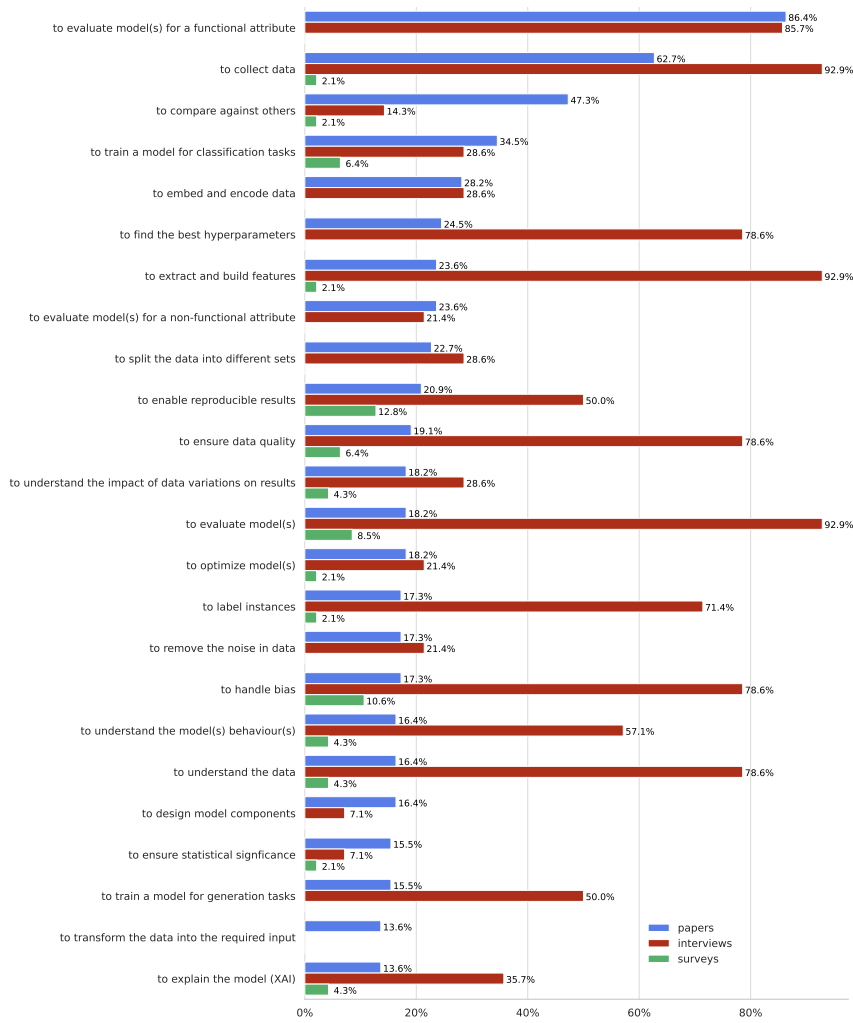


**Fig. 9** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *purpose/output category* appears. The figure displays approximately the first third of the *purpose/output codes*, organized by the subject *papers* appearance.

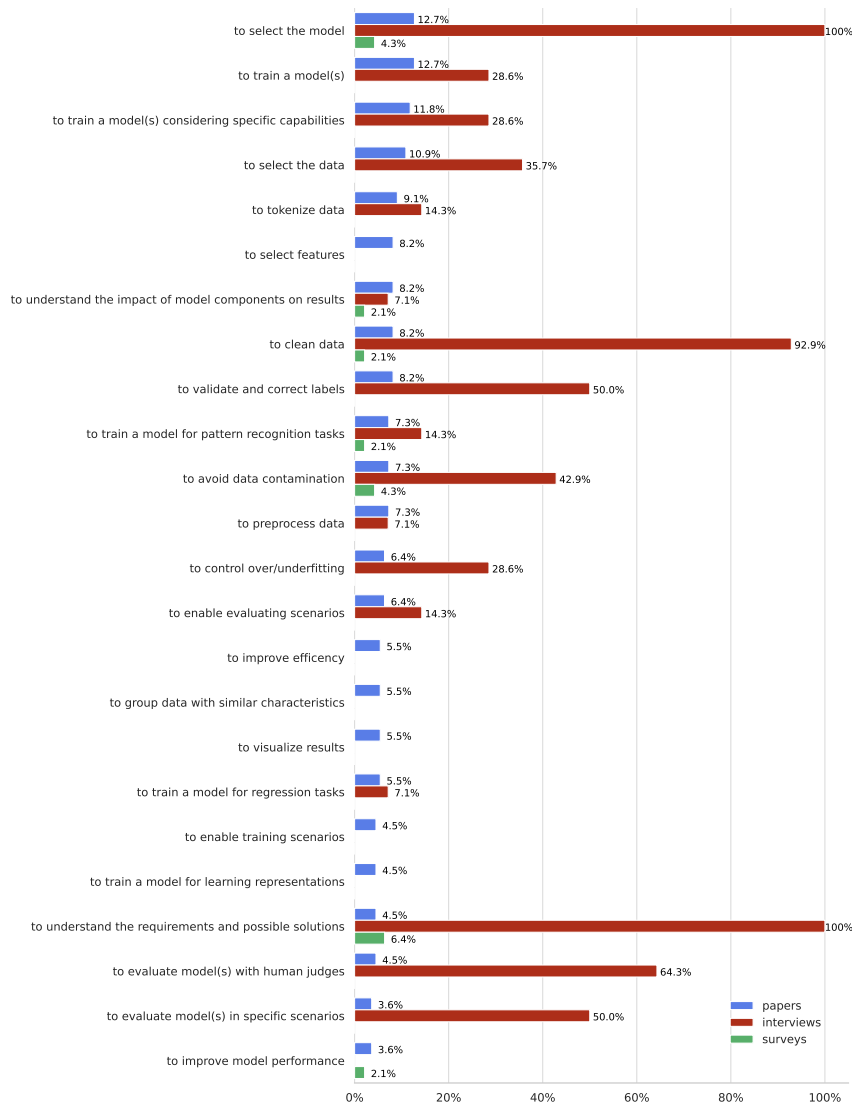


**Fig. 10** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *purpose/output category* appears. The figure displays approximately the second third of the *purpose/output codes*, organized by the subject *papers* appearance.

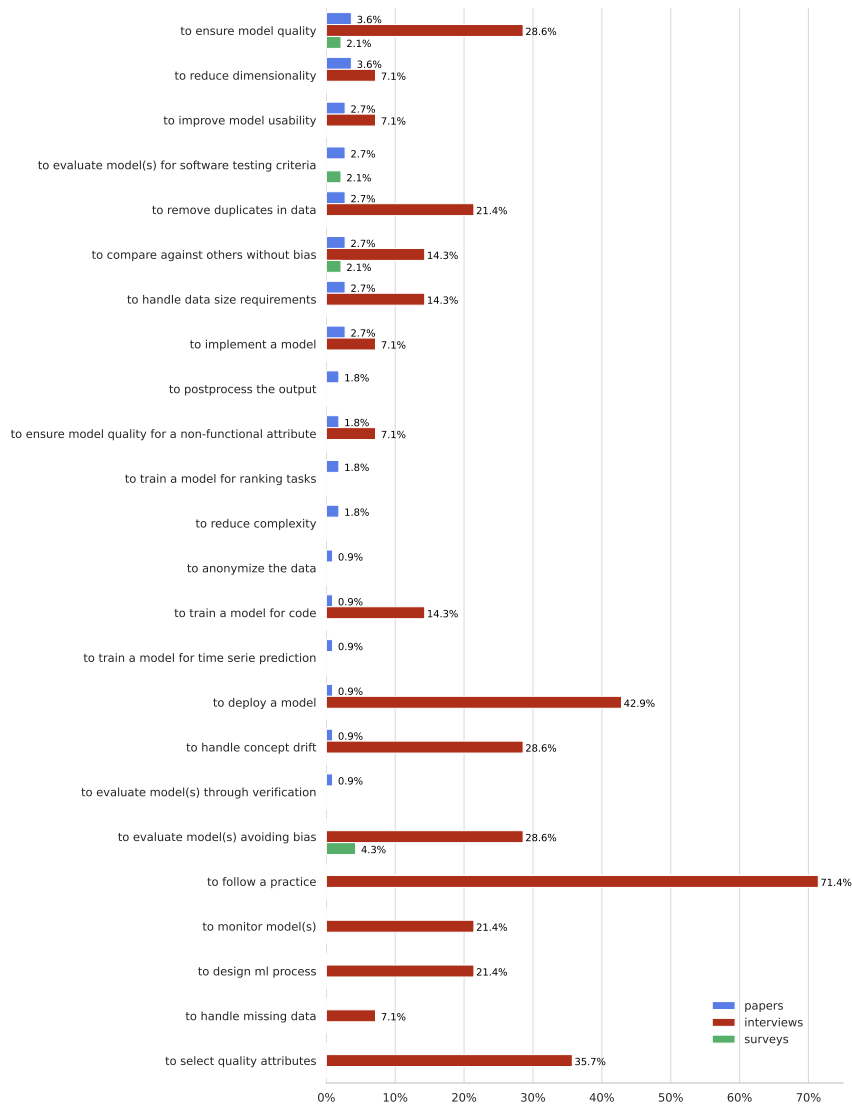


**Fig. 11** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *purpose/output category* appears. The figure displays approximately the last third of the *purpose/output codes*, organized by the subject *papers* appearance.

## C.2 Categories associated to the ML Stage

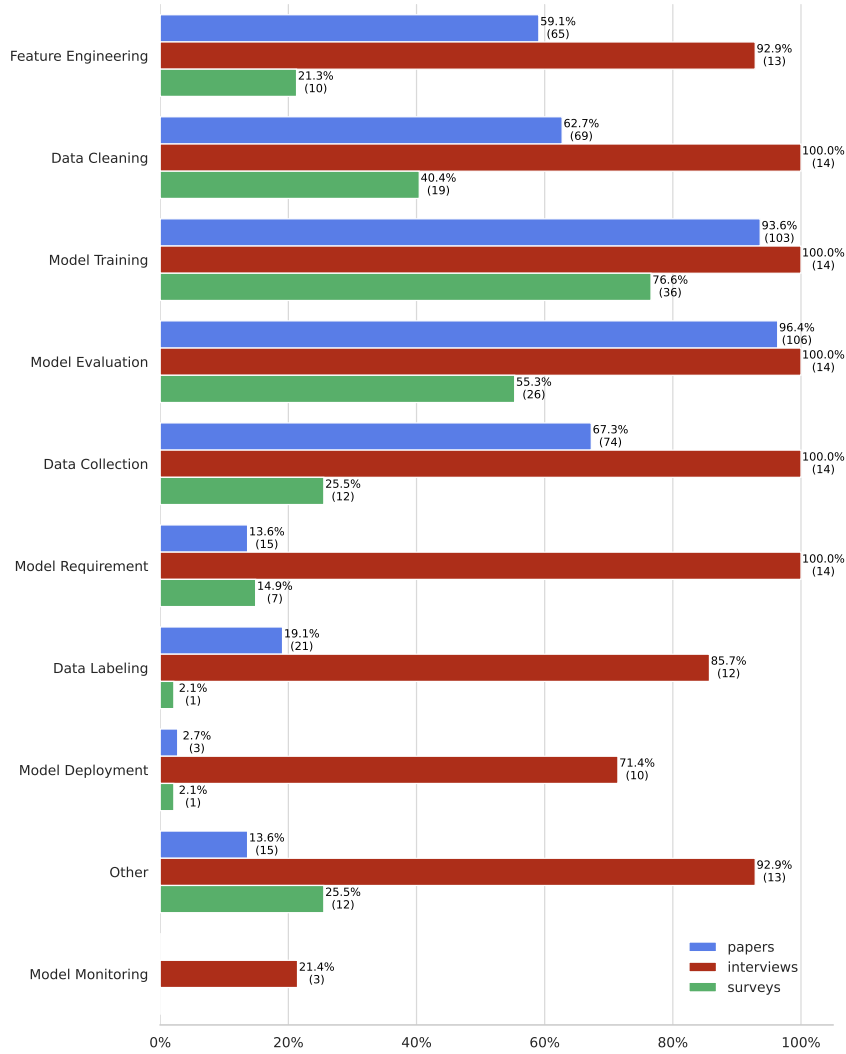


**Fig. 12** Percentage of subjects (*i.e.*, papers, interviewees, and surveys) in which each *ML stage category* appears.

---

### C.3 Categories associated to the SE Tasks

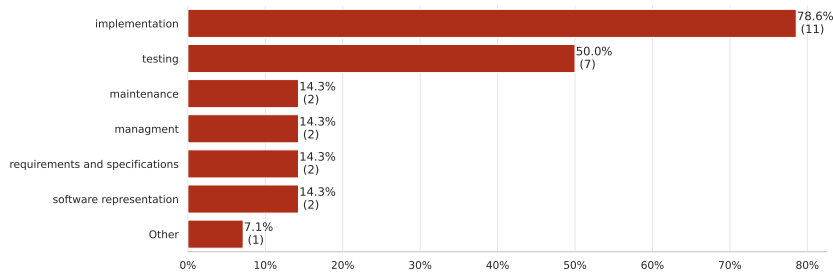


**Fig. 13** Percentage of interviews in which each *SE task category* appears.

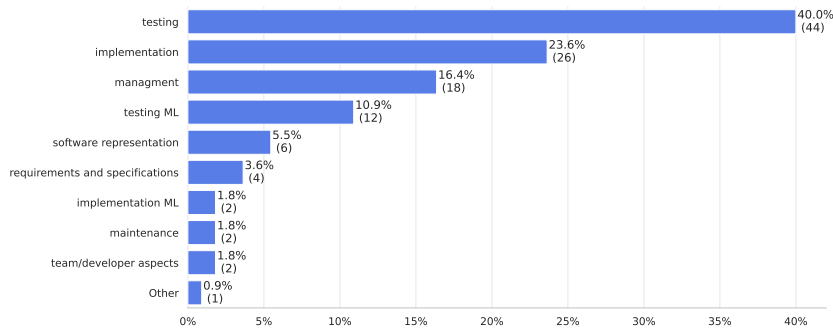


**Fig. 14** Percentage of papers in which each *SE task category* appears.

## C.4 Codes associated to the Quality attributes

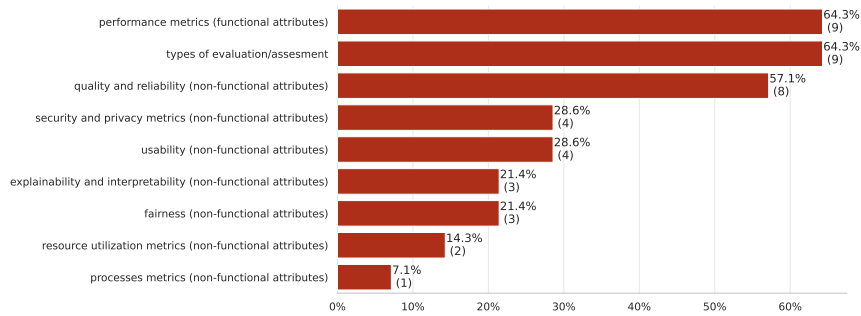


**Fig. 15** Percentage of interviews in which each *quality attribute category* appears.

## C.5 Codes associated to the ML Challenges

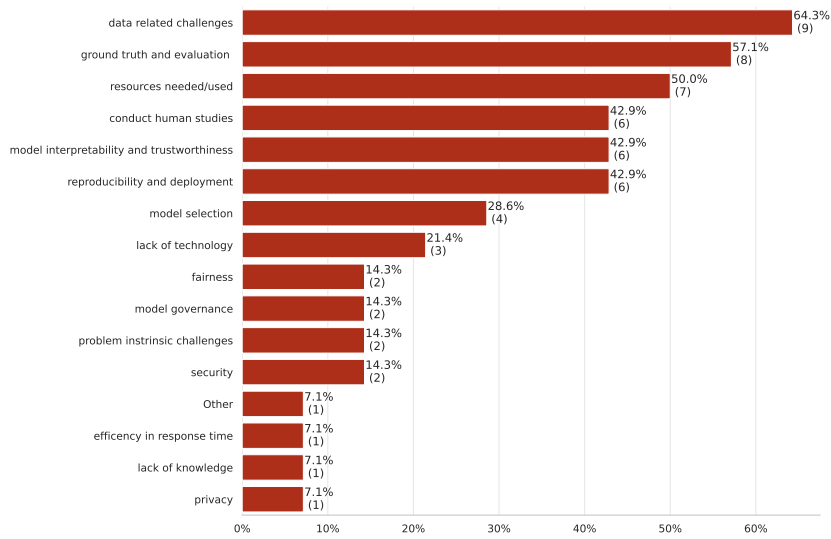


**Fig. 16** Percentage of interviews in which each *challenge category* appears.

## C.6 Codes associated to the Reviewer's perspective

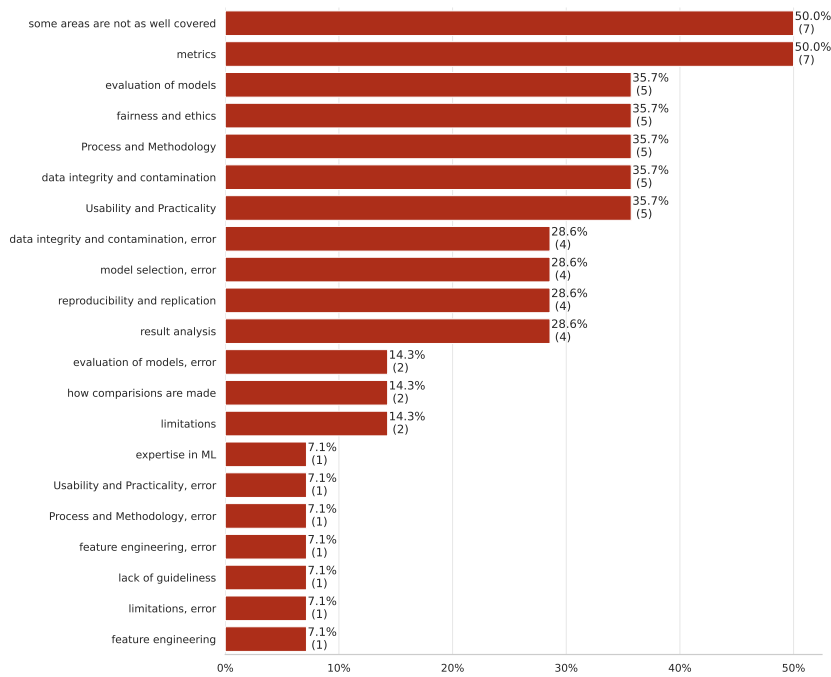


**Fig. 17** Percentage of interviews in which each *reviewers' perspective category* appears.

---

## C.7 Codes associated to the Educator's perspective

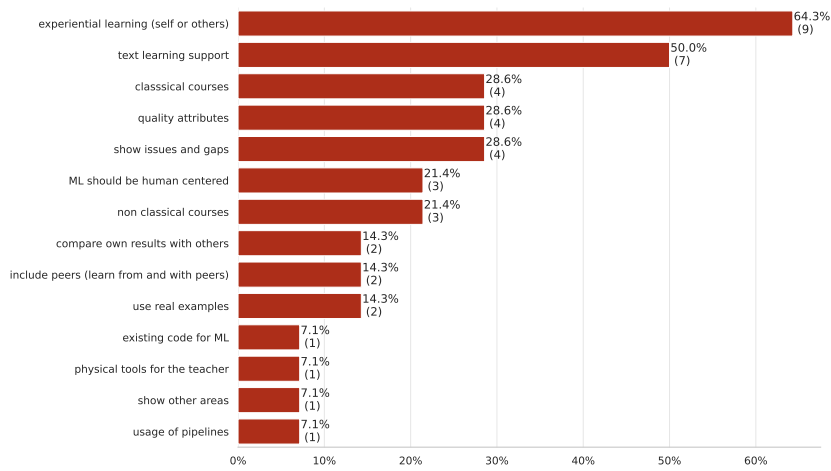


**Fig. 18** Percentage of interviews in which each *educators' perspective category* appears.

## D Supplementary Material for the Results of the Open Coding, 2023-2025

### D.1 Codes associated to the ML practices, 2023-2025

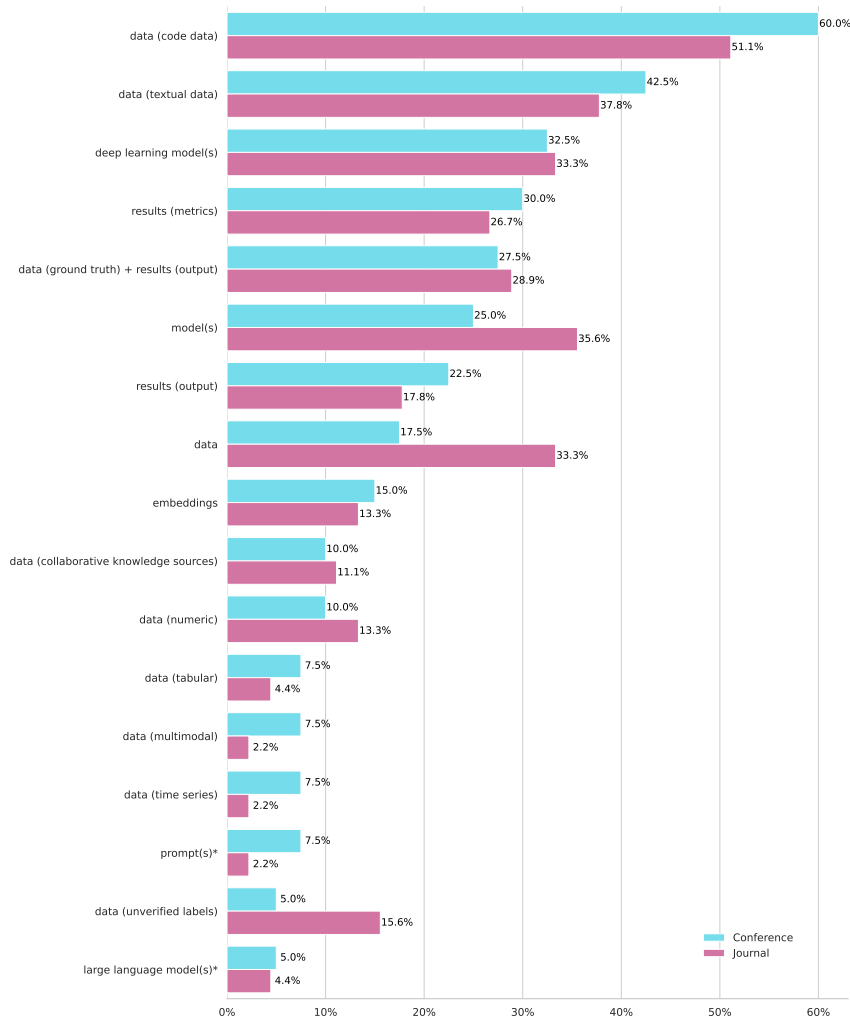


**Fig. 19** Percentage per venue (*i.e.*, conferences, journals) in which each *input category* appears. The figure displays the first half of the *input codes*, organized by the subject *papers* appearance.

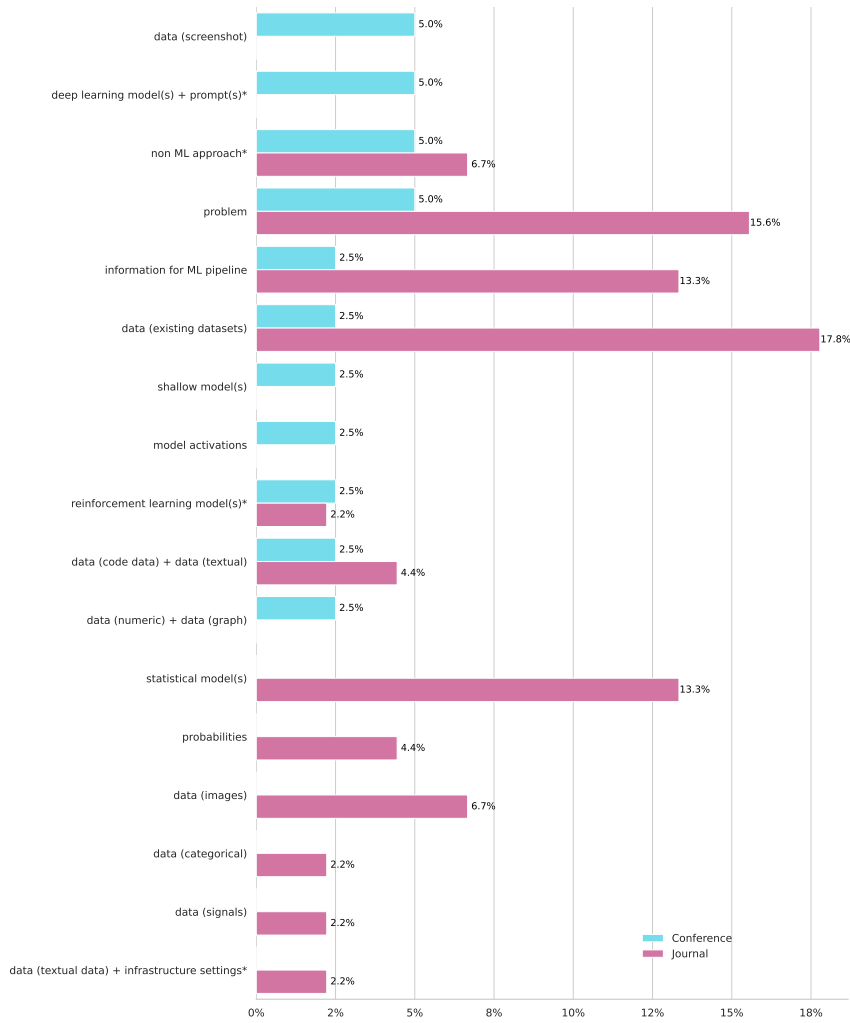


**Fig. 20** Percentage per venue (*i.e.*, conferences, journals) in which each *input category* appears. The figure displays the second half of the *input codes*, organized by the subject *papers* appearance.

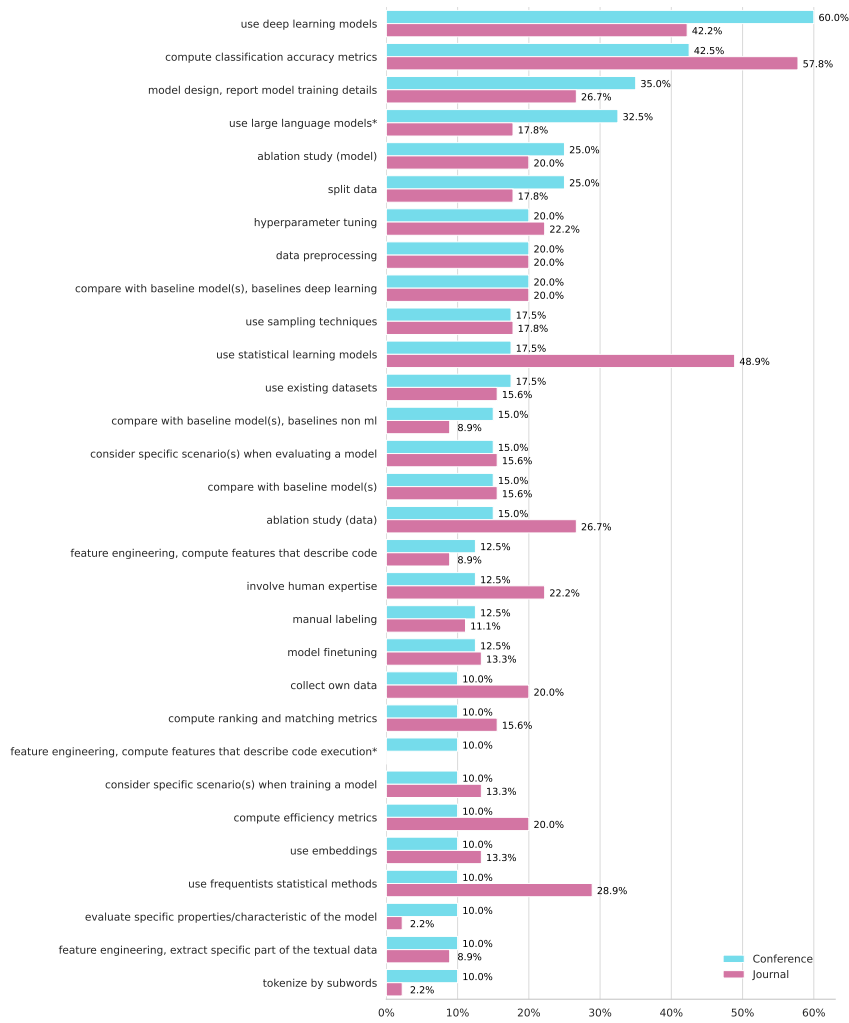


**Fig. 21** Percentage per venue (*i.e.*, conferences, journals) in which each *technique category* appears. The figure displays approximately the first fifth of the *technique codes*, organized by the subject *papers* appearance.

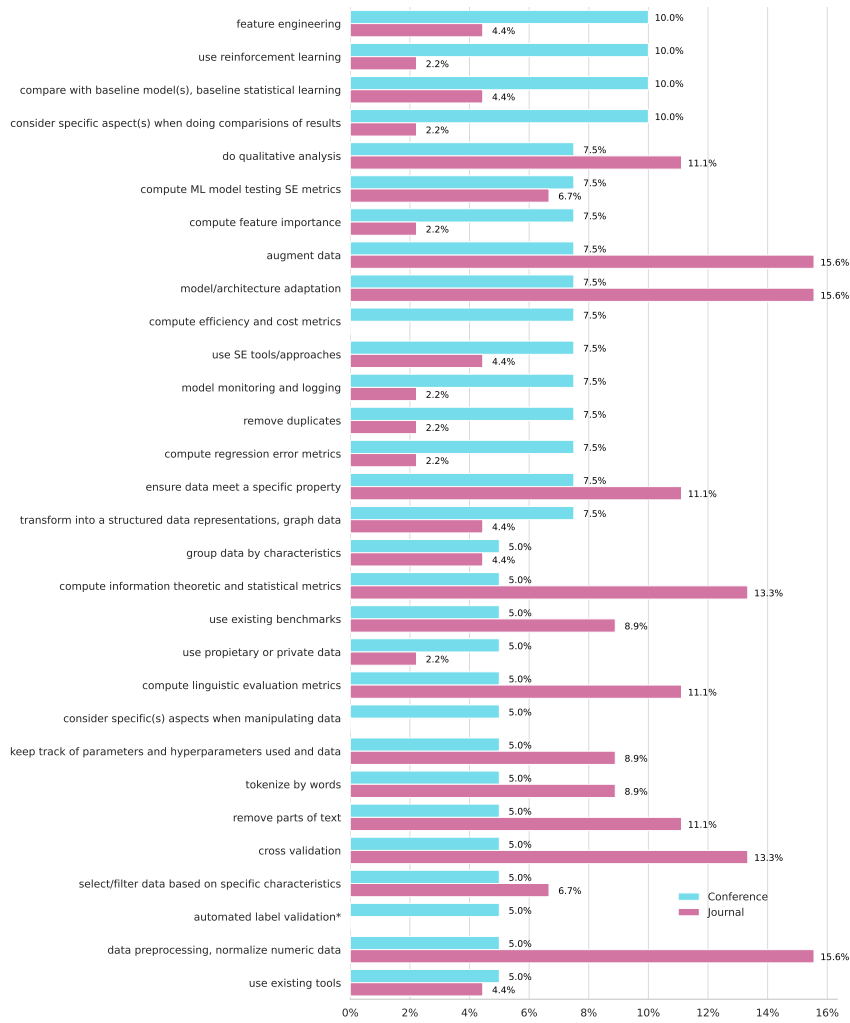


**Fig. 22** Percentage per venue (*i.e.*, conferences, journals) in which each *technique category* appears. The figure displays approximately the second fifth of the *technique codes*, organized by the subject *papers* appearance.

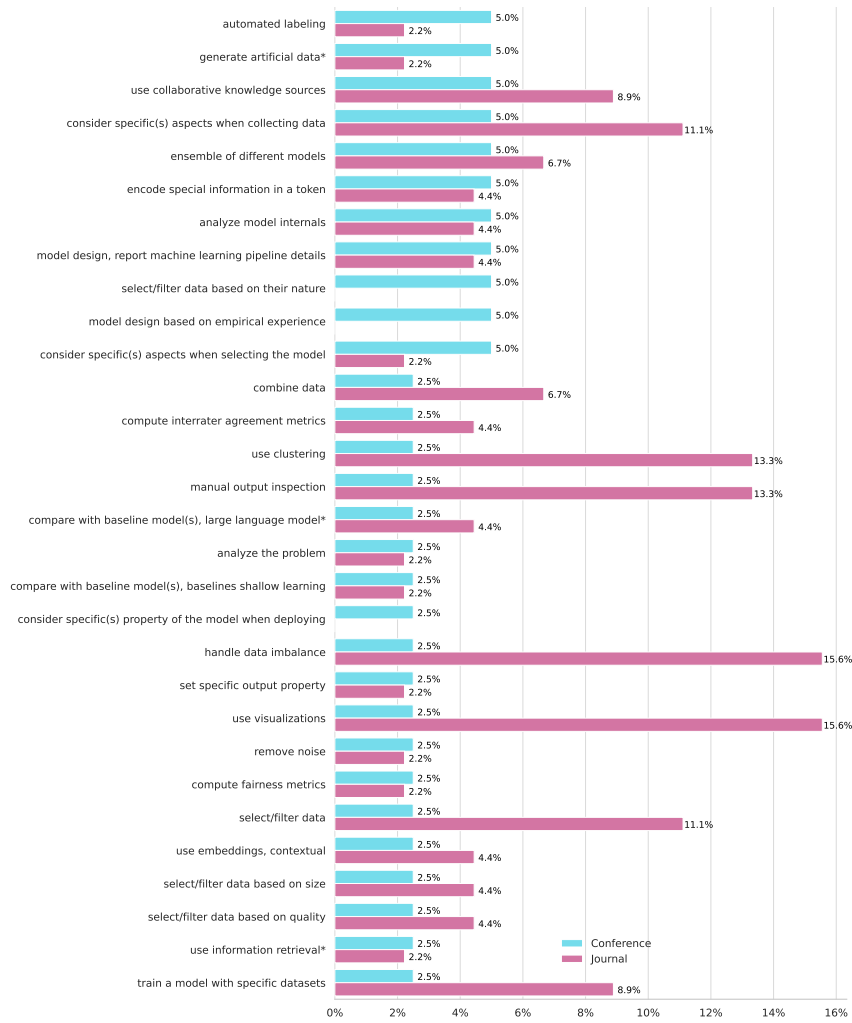


**Fig. 23** Percentage per venue (*i.e.*, conferences, journals) in which each *technique category* appears. The figure displays approximately the third fifth of the *technique codes*, organized by the subject *papers* appearance.

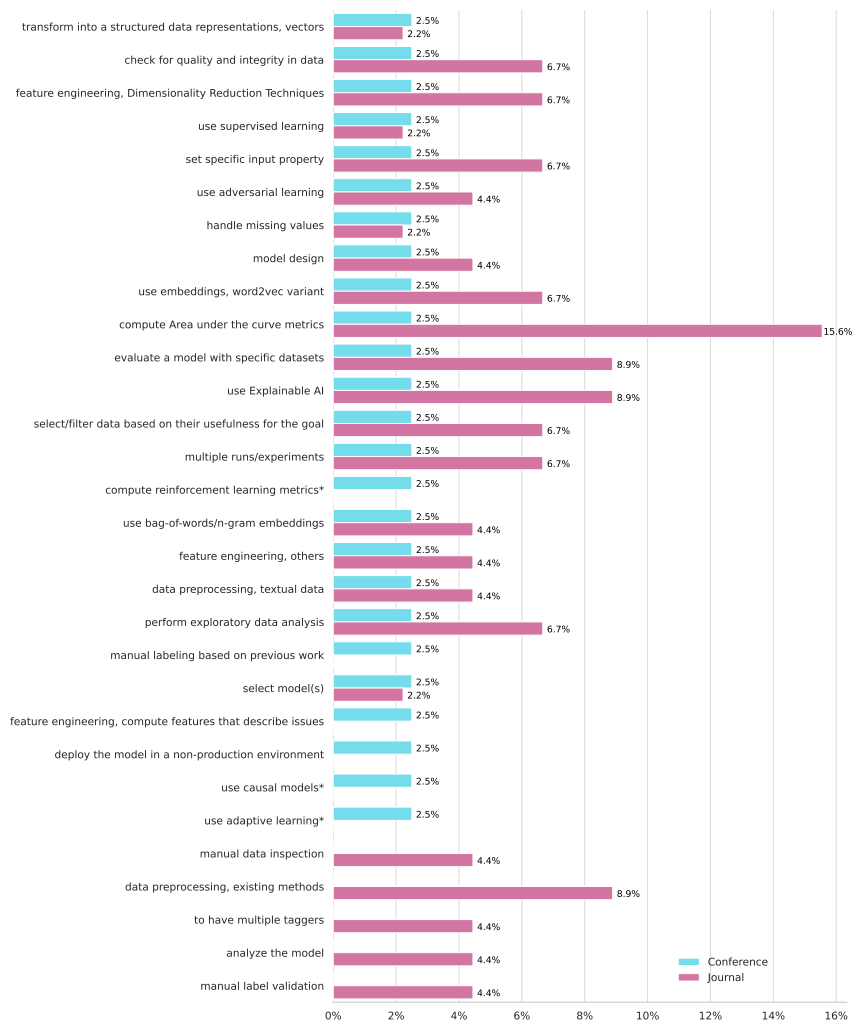


**Fig. 24** Percentage per venue (*i.e.*, conferences, journals) in which each *technique category* appears. The figure displays approximately the fourth fifth of the *technique codes*, organized by the subject *papers* appearance.

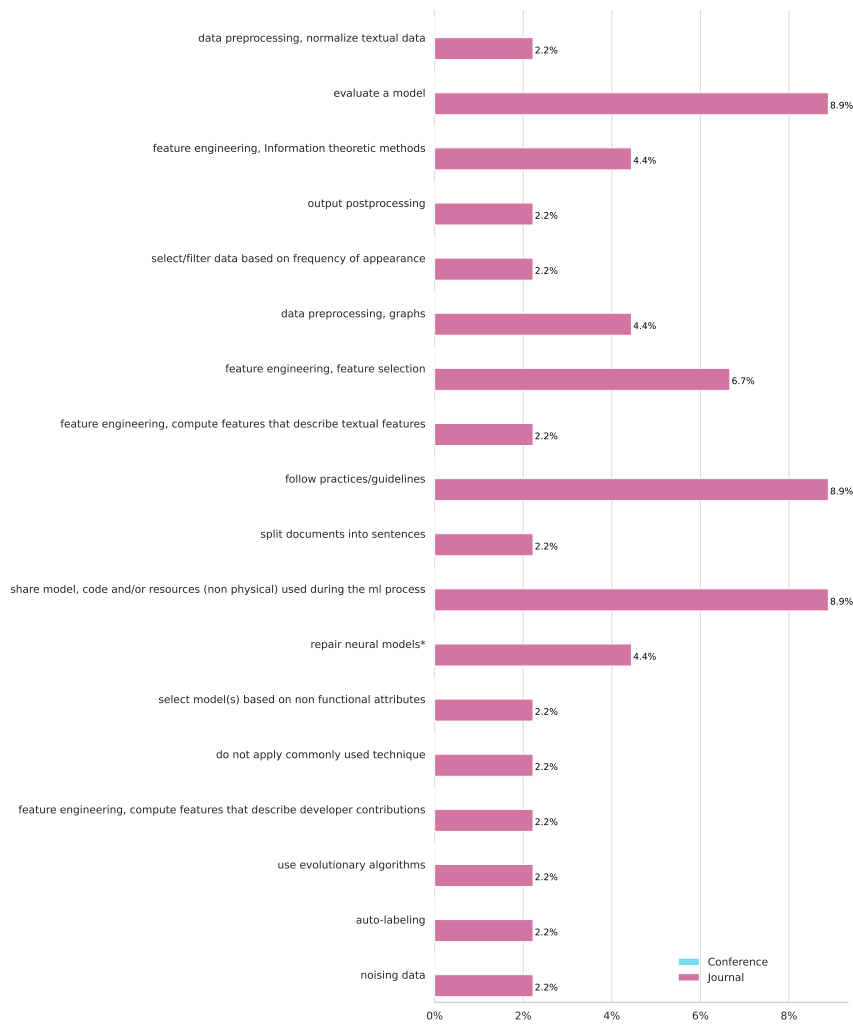


**Fig. 25** Percentage per venue (*i.e.*, conferences, journals) in which each *technique category* appears. The figure displays approximately the last fifth of the *technique codes*, organized by the subject *papers* appearance.

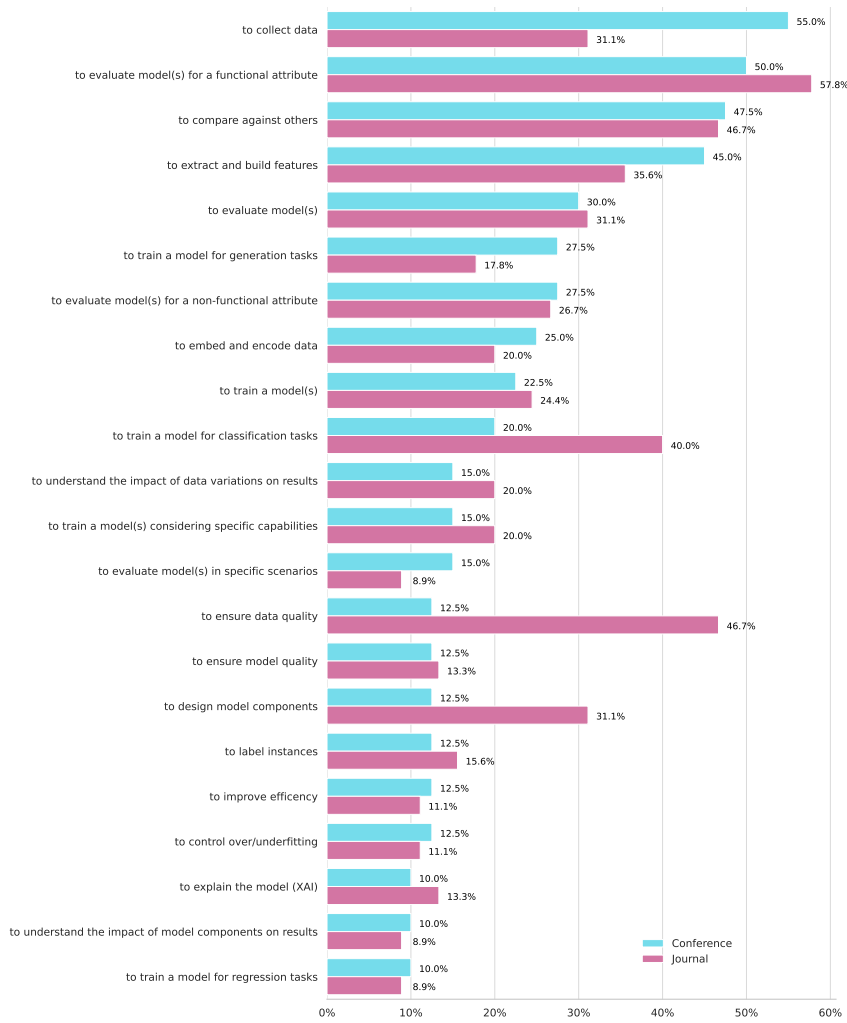


**Fig. 26** Percentage per venue (*i.e.*, conferences, journals) in which each *purpose/output category* appears. The figure displays approximately the first third of the *purpose/output codes*, organized by the subject *papers* appearance.

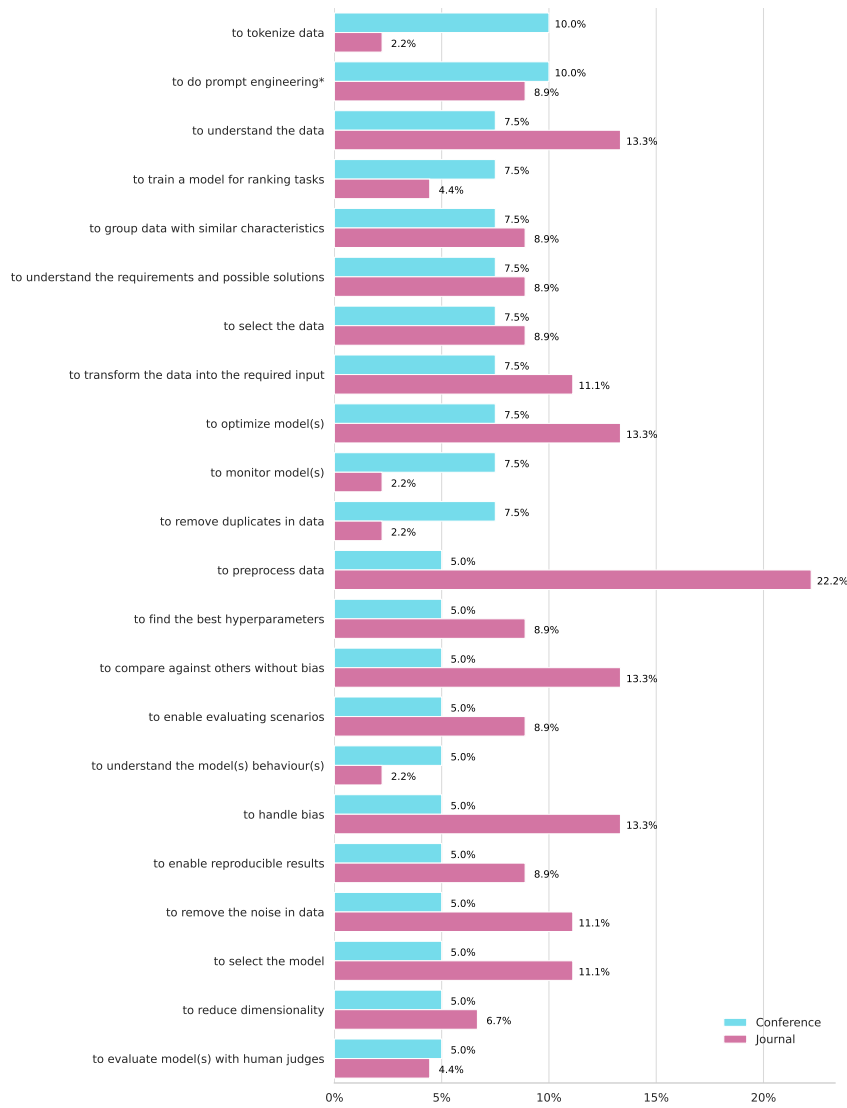


**Fig. 27** Percentage per venue (*i.e.*, conferences, journals) in which each *purpose/output category* appears. The figure displays approximately the second third of the *purpose/output codes*, organized by the subject *papers* appearance.

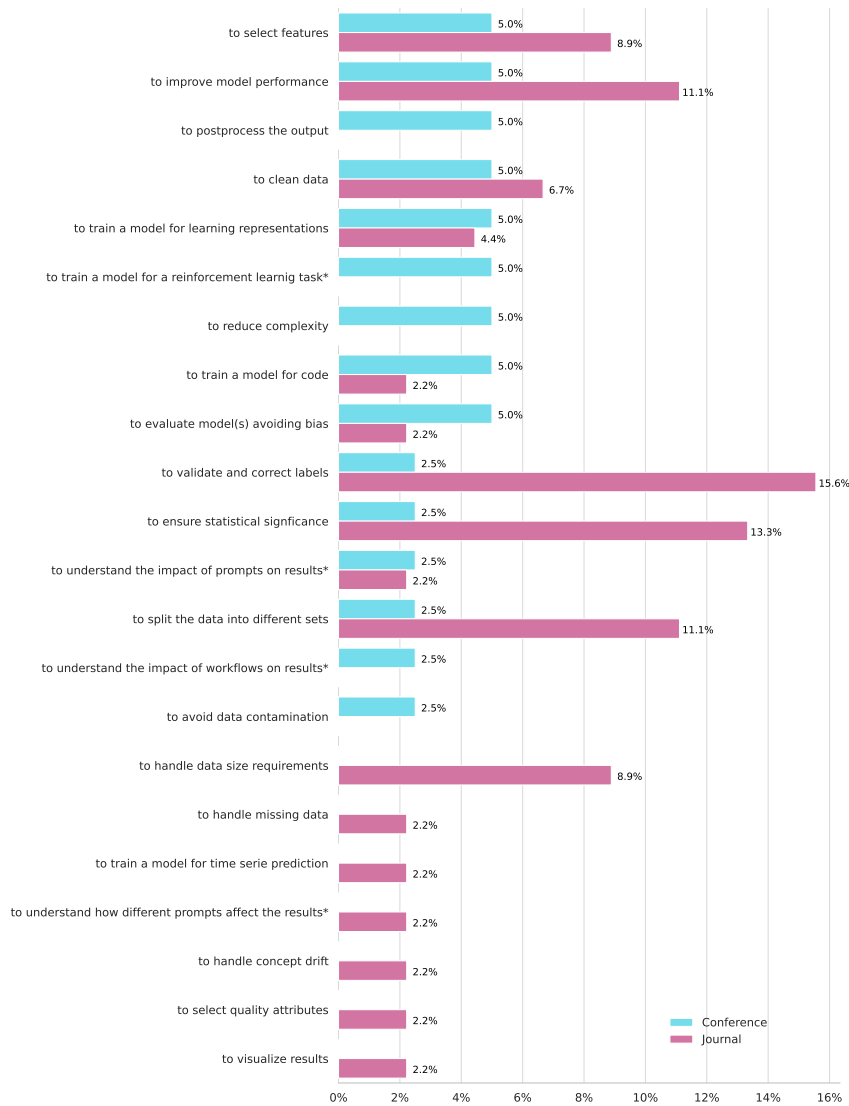


**Fig. 28** Percentage per venue (*i.e.*, conferences, journals) in which each *purpose/output category* appears. The figure displays approximately the last third of the *purpose/output codes*, organized by the subject *papers* appearance.

## D.2 Categories associated to the ML Stage, 2023-2025

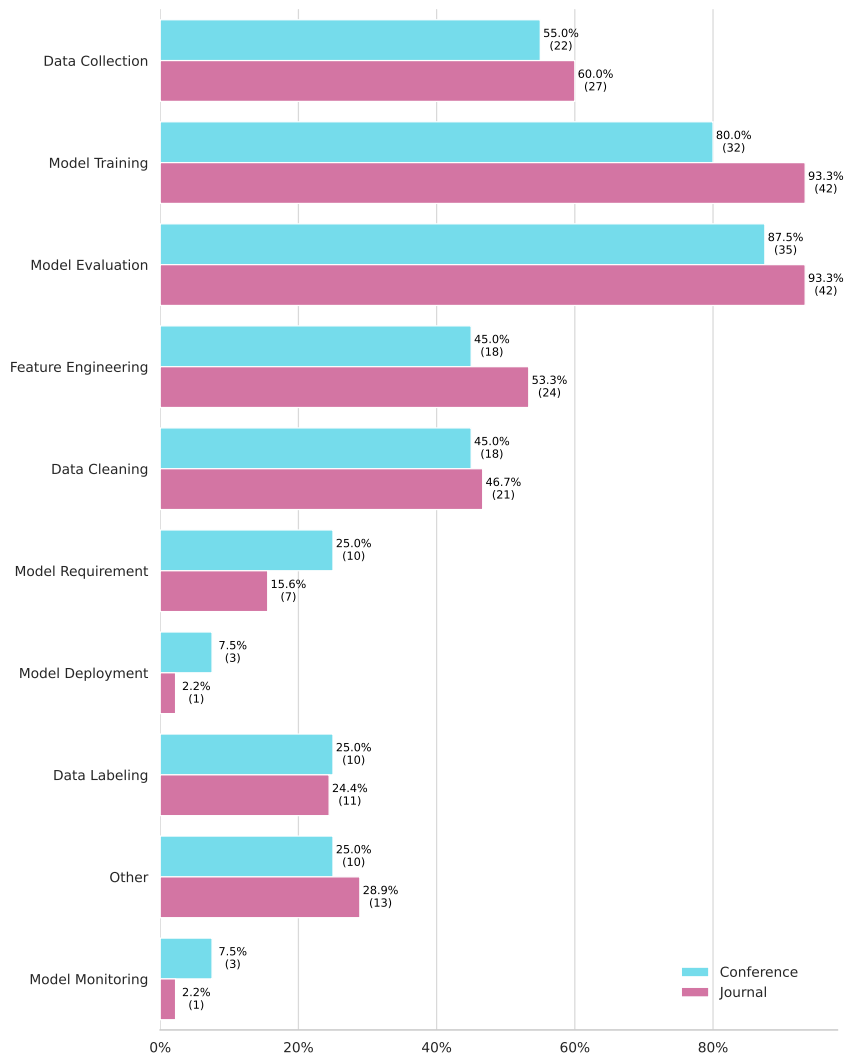


**Fig. 29** Percentage per venue (*i.e.*, conferences, journals) in which each *ML stage category* appears.

### D.3 Categories associated to the SE Tasks, 2023-2025

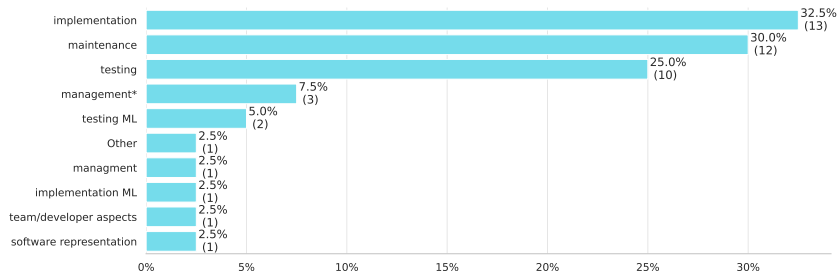


**Fig. 30** Percentage of journals in which each *SE task category* appears.

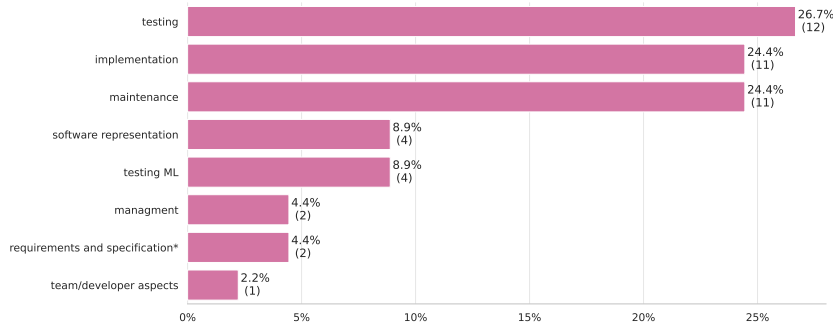


**Fig. 31** Percentage of conferences in which each *SE task category* appears.