Terms

AI-ML-based systems:
Systems that incorporate at least one component with artificial intelligence (AI), e.g. one with a machine learning (ML) model.

Studies on the state of practice of software engineering:
Studies that describe or analyze software engineering activities (e.g. requirements engineering, design, development, etc.) in a professional context (usually industry). They can be published by industry practitioners or by external academic researchers. Examples of such studies are case studies, surveys, interviews, grey literature reviews, or experience reports. We explicitly exclude industry papers that focus on the usage of or experience with a very narrow scope, i.e. a specific technology or approach.

The primary studies we want to identify with this SLR are therefore state of practice studies in the context of software engineering for AI-ML-based systems. In this case, “AI-ML-based systems” means that there must be an explicit system focus, in contrast to a focus on AI/ML models or algorithms as is common in data science.

Research Questions

1. What studies exist on the state of practice for software engineering for AI-ML-based systems?
    a. What methods did these studies use?
    b. To which domains did the reported systems belong?
    c. What types of ML/AI were reported in these studies?
    d. What were the characteristics of the study participants?
2. What challenges for the development of AI-ML-based systems do software professionals experience?
3. What are effective practices for the development of AI-ML-based systems as reported by software professionals?

Search Engines and Digital Libraries

Google Scholar (https://scholar.google.com)
ACM (https://dl.acm.org/)
Science Direct (https://www.sciencedirect.com/)
World Wide Science (https://worldwidescience.org/)
Web of Knowledge (https://webofknowledge.com)
IEEE Xplore (https://ieeexplore.ieee.org/)

Search Strategy

Get a startset for snowballing by using a search string in the different search engines and digital libraries. Here the filters will be set. The timespan is chosen like this to get a better reproducibility. With this set as a basis, the next step will be snowballing by Wohlin (https://dl.acm.org/doi/10.1145/2601248.2601268). The citations will be searched by using Google Scholar. This is also the default search engine for the references in the used papers. If an interesting reference cannot be found by this search engine, it will be marked by a notice how it was found. A paper would be excluded immediately, when the title includes "artificial intelligence in/for software engineering" or "machine learning in/for software engineering" because it is obviously the wrong direction. Also it will be excluded when the abstract makes sure that the paper is not about software engineering in/for AI/ML-based systems and also if the text didn't include information about software engineering in/for AI/ML-based systems in practice. This means that if only one of these things occur the paper will be immediately excluded. For all papers from the snowballing phases the first publication that mentions this paper will be marked as the source/origin.

For World Wide Science all papers that will be shown by using the search string were look at, but only the papers that includes "machine learning" and "software engineering" in the title of the Document will be listed in the sheet. It seems like the search engine ignores a part of the search string even if it is in the format that the help site explains.
Choose a maximum of 250 Papers for each forward snowballing round as stop criterion for the snowballing.

Searchstring

- allintitle: machine learning software engineering
- Adoption (if allintitle didn't exist):
  Title: "machine learning" AND "software engineering"

Filters

- Years: 2010-2019
- Language: English

Inclusion Criteria

1. The paper is a software engineering state of practice study as defined above
2. The paper is in the context of one or several AI-ML-based systems as defined above.

Exclusion Criteria

1. The paper was not published between 2010 and 2019 (last AI Wave).
2. The paper is not in English.
3. The paper is not a scientific peer-reviewed publication.
4. The full text of the paper is not available to the author.

5. The paper is a duplicate or extension of an already included paper. For duplicates, the older one is kept. For extensions, the newer one is kept.
6. The paper is a secondary or tertiary study (e.g. SLR or SMS).

Results

SLR Phase	Description	Number new found Articles / Total Number used Articles
Applying Search String to Retrieve Start Set	Create start set by entering the search string in every search engine and digital library from above	Totally found: 186
Deduplication	Remove duplicates from start set	Articles left:136
Inclusion/Exclusion	Perform filtering with inclusion and exclusion criteria	Articles used: 8
Backward and Forward Snowballing of Start Set	Create a new Set (second set) with the referenced literature and the literature that cite the literature from the start set. Then Filter this set with inclusion and exclusion criteria.	19 /27
Backward and Forward Snowballing of Second Set	Create a new Set (third set) with the referenced literature and the literature that cite the literature from the second set. Then Filter this set with inclusion and exclusion criteria.	22 /49
Snowballing of Third set (Forward and Backward)	Create a new Set (fourth set) with the referenced literature and the literature that cite the literature from the third set. Then Filter this set with inclusion and exclusion criteria.	6 /55
Writing Phase	Write the found information into a Bachelor thesis	0 /55

Result Tables

At the end it will be one table for the literature and one table for the interesting information: https://github.com/st140937/bachelorThesis

Artifacts

First phase: table with all found papers (including marked duplicates)
For all other phases: a table with all relevant papers (methodology of snowballing)
At the end a table with all relevant information for the RQ.
After every Phase a new combination of iteration and snowballing direction will be created.