What Is an AI Engineer? An Empirical Analysis of Job Ads in The Netherlands

Marcel Meesters
m.meesters@fontys.nl
Fontys University of Applied Science
Eindhoven, Netherlands

Petra Heck
p.heck@fontys.nl
Fontys University of Applied Science
Eindhoven, Netherlands

Alexander Serebrenik
a.serebrenik@tue.nl
Eindhoven University of Technology
Eindhoven, Netherlands

ABSTRACT

Recently, the job market for Artificial Intelligence (AI) engineers has exploded. Since the role of AI engineer is relatively new, limited research has been done on the requirements as set by the industry. Moreover, the definition of an AI engineer is less established than for a data scientist or a software engineer. In this study we explore, based on job ads, the requirements from the job market for the position of AI engineer in The Netherlands. We retrieved job ad data between April 2018 and April 2021 from a large job ad database, Jobfeed from TextKernel. The job ads were selected with a process similar to the selection of primary studies in a literature review. We characterize the 367 resulting job ads based on meta-data such as publication date, industry/sector, educational background and job titles. To answer our research questions we have further coded 125 job ads manually.

The job tasks of AI engineers are concentrated in five categories: business understanding, data engineering, modeling, software development and operations engineering. Companies ask for AI engineers with different profiles: 1) data science engineer with focus on modeling, 2) AI software engineer with focus on software development, 3) generalist AI engineer with focus on both models and software. Furthermore, we present the tools and technologies mentioned in the selected job ads, and the soft skills.

Our research helps to understand the expectations companies have for professionals building AI-enabled systems. Understanding these expectations is crucial both for prospective AI engineers and educational institutions in charge of training those prospective engineers. Our research also helps to better define the profession of AI engineering. We do this by proposing an extended AI engineering life-cycle that includes a business understanding phase.

KEYWORDS

data science, software engineering, AI engineer, ML engineer, job market, job ad

ACM Reference Format:

Marcel Meesters, Petra Heck, and Alexander Serebrenik. 2022. What Is an AI Engineer? An Empirical Analysis of Job Ads in The Netherlands. In *1st Conference on AI Engineering - Software Engineering for AI (CAIN'22), May 16–24, 2021, Pittsburgh, PA, USA.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3522664.3528594

CAIN'22, May 16–24, 2022, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-9275-4/22/05…$15.00
https://doi.org/10.1145/3522664.3528594

1 INTRODUCTION

AI has gained increasing interest and adoption in the IT industry [2]. Professional development and deployment of software systems that incorporate AI features is challenging and requires skilled engineers [1]. This software engineering for AI-enabled systems is also called "AI engineering" [2]. A 2018 report from Indeed [22] found ML[1] engineer to be the "most pure AI job".

AI engineering is growing as a separate discipline from both software engineering [4] and data science [7, 15]. Many universities offer bachelor and master degrees in AI engineering, and more and more companies offer positions for software engineers or data scientists to build AI-enabled software systems. Hence, to support curriculum designers and teachers of AI engineering programs we need a more thorough understanding of **what companies require from AI engineers**.

As Daneva et al. [8] argue, job ads offer a cost-effective way to study a broad range of companies. Inspired by the work of Daneva et al. on the profession of requirements engineer, we decided to use job ads as a starting point also for our study on AI engineers. We formulated the following research questions:

- RQ1a: Which job tasks are expected from AI engineers?
- RQ1b: Are the job tasks expected from AI engineers more focused on data science or on software engineering?
- RQ2: What are the most frequently mentioned tools and technologies expected to be used by AI engineers?
- RQ3: Which soft skills are required for the job of AI engineer?

In our study we investigated what companies require from AI engineers by studying 367 job ads. We conclude that AI engineers are requested under a diversity of job titles. They perform job tasks in five categories: business understanding, data engineering, modeling, software development and operations. In relation to the division between data science job tasks and software engineering job tasks we observe three distinct profiles of AI engineers. We also report on tools, technologies and soft skills requested in the job ads. Programming languages like Python and AI tools like Tensorflow are wide-spread, but more specific tooling intended to support AI engineering like MLFlow is hardly found in the job ads. Learning ability and team orientation are the most requested soft skills.

[1]With the current state of the practice many AI systems involve machine learning (ML). Although ML is in fact a sub field of AI, in the context of this paper we consider these two terms synonyms.

Our research helps to understand the expectations companies have for professionals working with AI-enabled systems. Understanding these expectations is crucial both for prospective AI engineers and educational institutions in charge of training those prospective engineers.

In the remainder of this paper we first introduce background and related work on AI engineering and job ad research. Section 3 explains the job ad selection and coding process. Section 4 contains meta-data about the selected job ads. Section 5 presents the answers to the research questions. Section 6 discusses the implications of our findings for AI engineering research, education and industry. Section 7 analyzes threats to validity. The conclusion summarizes our contributions and presents future work.

2 BACKGROUND AND RELATED WORK

In this section we present background and related work on AI engineering, the AI engineer, job ad research and the way education level is specified in job ads in The Netherlands.

2.1 AI Engineering

In literature, different definitions of AI engineering can be found. We did not find one single definition that is generally accepted as *the* definition of AI engineering. In this section we sum up the different existing definitions in literature. Based on those definitions we extract our working definition of AI engineering that we use for selecting AI engineer job ads.

Bosch et al. [2] define AI engineering as "a set of methods and tools that originated from software engineering in a system life cycle, and procedures, technologies and tools from data science and AI". Burkov [3] defines ML engineering as "the use of scientific principles, tools, and techniques of machine learning and traditional software engineering to design and build complex computing systems". A notion related to ML engineering is MLOps: "an ML engineering culture and practice that aims at unifying ML system development (Dev) and ML system operation (Ops)." [10] Farah [9] defines MLOps by depicting the essential steps in an MLOps process, see Figure 1. Lwakatare et al. [20] present a similar picture in their paper on DevOps for AI. They observe that, in practice, the four phases of Farah are not necessarily followed sequentially and many more feedback loops are involved.

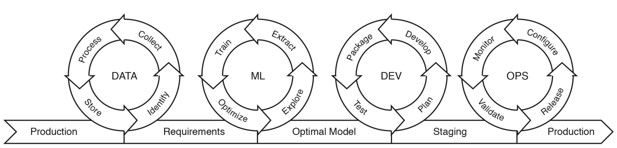

Figure 1: The AI engineering life cycle [9]

Another definition is that of Sato et al. [26] who have proposed Continuous Delivery for ML (CD4ML), "a software engineering approach in which a cross-functional team produces ML applications based on code, data, and models in small and safe increments that can be reproduced and reliably released at any time, in short adaptation cycles."

Yet another term used in this context is "Software engineering for AI/ML" [4] [1].

Although the above definitions are all different, they have common elements. They all contain machine learning aspects and (software) engineering aspects. Next to that they aim at releasing AI-enabled software systems in a production environment. From these common elements, we build the following definition, that we use for selecting AI engineer job ads:

> *AI Engineering is a combination of machine learning and software engineering with the goal to build production-ready machine learning systems.*

2.2 The AI Engineer

To our knowledge there are no academic publications on the profession of AI engineer as such. Somewhat related is the work of Kim et al. [15] who surveyed 793 professional data scientists in software engineering teams at Microsoft. The authors identified 9 distinct clusters of data scientists, and their corresponding characteristics. This survey was conducted from the data scientist perspective where we also take the software engineering perspective into account. On the one hand, Kim et al. had much more detailed data in the survey than we have in the job ads. On the other hand, they present the context of only one albeit very large company, where our analysis includes many different companies from different industries.

We found two industry reports that analyze the job market for AI in general, not specifically for AI engineers.

In 2018, Indeed [22], published a short report analysing AI jobs in the US market. The report lists the top-ten job titles requiring AI and ML skills, with ML engineer ranked first. In our analysis we explicitly focus on the job ads for ML engineer for The Netherlands and report a list of job titles for ML engineer positions.

In 2019, LinkedIn [19], brought out a report on the dynamics between AI and the European labour market. As the basis for this report the authors used keywords and a ML model to select a set of LinkedIn profiles as being "AI talent". The LinkedIn report also lists skills (programming languages, AI-related libraries of code, data science libraries, and soft skills) requested in job ads on LinkedIn. This is similar to our analysis for RQ2 and RQ3, but we focus on AI engineers in The Netherlands and have a more recent time frame. In addition our analysis includes a detailed manual coding of job tasks from the job ads (RQ1). The LinkedIn report does not provide any information on such job tasks, nor on job titles.

No industry reports on job tasks for AI engineers (RQ1) were found.

2.3 Job Ad Research

Job ads have been often used in occupational research as they are a representation of the job characteristics and terminology [14, 24]. Although research on job requirements for AI engineers is new, software engineering job ads have been studied in the past [5, 8, 23, 27]. For example, Daneva et al. [8] have investigated the background required from requirements engineers, and characteristics of their jobs in terms of competences and responsibilities. Another example is the study of the testing profession by Cerioli et al. The authors have observed that six times more testers are sought than coders, and that while unit testing is the most required skill for coders, acceptance testing is the most popular for testers [5].

2.4 Required Education Level in Job Ads in The Netherlands

Job ads in The Netherlands refer to the education level they require by specifying the type of university: University of Applied Science (UAS) or General University (GU). As the name indicates, a degree from UAS is mostly application-oriented, while a degree at a GU is more scientifically oriented. Job ads may also specify a PhD as required educational background. In this paper we thus use the distinction UAS, GU or PhD when we report on the required education level in the job ads.

3 JOB AD SELECTION AND CODING

Our approach to selecting job ads to answer the research questions is inspired by the process of selecting primary studies in a systematic literature review [16]:

(1) Define inclusion and exclusion criteria
(2) Design query string
(3) Identify databases and other sources to search
(4) Select relevant job ads based on full job ad
(5) Classify resulting job ads

As a final step, we coded [25] the resulting job ads to answer the research questions. Each step was executed by the first author and checked by the second author. The outcome of the step was adjusted according to the discussion between the two authors. The following paragraphs describe each of the steps in detail. The classification step is described in Section 4.

3.1 Inclusion and Exclusion Criteria

We use two *inclusion* criteria. First, the job ad is for an AI engineer, i.e. a combination of machine learning and software engineering with the goal to build production-ready machine learning systems. Second, to ensure that the answers to our research questions reflect the current company requirements we included job ads from the last three years, *i.e.,* from April 2018 till April 2021 (data collection period). We *exclude* job ads with an empty function description because we need this description to extract the answers to the RQs.

3.2 Query String

In Section 2.1 we defined AI engineering as the combination of software engineering and machine learning to build production-ready machine learning systems. To identify jobs related to AI engineering, we searched for job ads that contained both parts of the definition. We included developer as a synonym for software engineer. We included deep learning (DL) as a sub-field of ML.

(developer || engineer) & (AI || ML || DL)

We did not only include the abbreviations, but also the full terms "artificial intelligence", "machine learning" and "deep learning". Since companies advertising in The Netherlands might also publish job ads in Dutch, we also included the corresponding keywords in Dutch.

Table 1: Excluded job ads

	# Job Ads
No function description	182
Software Engineer	48
Data Engineer/Analyst	40
Other IT	20
Other Engineering	28
Lead/Product Owner/Manager	16
Architect	9
Scientific	5
Total excluded	348

3.3 Source Selection and Query Execution

While numerous websites allow companies to post job advertisements, many of those including such popular sites as Indeed or LinkedIn prohibit the use of their data for research purposes. Hence, as our data source we selected the Jobfeed database from TextKernel, a company that has allowed us to use their data. The Jobfeed-database contains over 1 billion current and historic job postings from websites from direct employers as well as from job-offering portals.

We executed the query on the entire job ad rather than on the job title only. After removing duplicates, we had a set of 715 unique job ads. Due to the conditions of our agreement with TextKernel we are not allowed to publicly share the job ads dataset.

3.4 Job Ad Selection

For the 715 unique job ads we determined whether the job ad is an AI engineer job ad that can be used for answering the research questions.

Note that 182 job ads had to be excluded because they merely contained a job title and no function description, associated tasks or responsibilities (see "No function description" in Table 1). These 182 job ads could, thus, not be used to answer our research questions.

The remaining 533 job ads where checked manually by the first author to determine if the function description does in fact describe an AI engineer job (a combination of machine learning and software engineering with the goal to build production-ready machine learning systems). The first author also performed card sorting with the job ads that were excluded, resulting in seven categories, see Table 1. In case of doubt the first author discussed with the second author until agreement was reached about inclusion or exclusion. The card sorting was used as a soundness check of this exclusion process (did we exclude for the right reason?).

In total, the first and the second authors agreed to exclude a total of 166 job ads that are not AI engineer job ads. These ads mention terms like AI or ML but are not focused on building production-ready machine learning systems.

The selection process resulted in 367 (= 715 - 348) AI engineer job ads.

Table 2: Job ads per Industry and Sector

Industry or Sector	# Job Ads
IT	50
Financial / Insurance	28
Trade	25
Business services	17
Healthcare / Welfare	7
Industry / Technology	7
Construction	6
Government / Non-profit	6
Media / Communication	6
Education / Research	4
Other / Unknown (e.g. agriculture, security, culture)	91
Intermediary Agency	116

3.5 Coding

To answer the research questions we first coded 100 randomly selected job ads out of the 367 job ads. Codes were developed using the initial coding approach [25]. After analyzing the first 100 coded job ads we randomly selected another 25 job ads from the remaining set and coded these. We compared the results of these 25 job ads with the initial coded set. No new codes had to be added for the additional set of job ads, from which we concluded that we achieved saturation. Therefore, we decided not to code any further jobs ads manually, meaning that the RQs were answered with a set of 125 manually coded job ads. After the initial coding, we applied axial coding [25], to come to categories for job tasks, technologies and soft skills.

4 JOB AD DATASET DESCRIPTION

In this section we characterize the 367 selected job ads based on meta-data such as publication date, industry/sector, educational background and job titles. This sets the ground for a more detailed analysis of the job characteristics in Section 5.

4.1 Publication Date

The resulting job ads are spread over the years 2018: 40, 2019: 98, 2020: 120, and 2021: 109 job ads. The data for 2018 and 2021 is incomplete as our dataset ranges from April 2018 till April 2021.

4.2 Industry and sectors

The dataset from the Jobfeed database contains a column "Industry/Sector" as meta-data. The advertised job positions come from various industries and sectors, as shown in Table 2. Not surprisingly, most ads are coming from the IT industry; financial institutions and trade are also popular. Also, 116 job positions are offered by intermediaries and agencies, and do not further specify industry or sector.

4.3 Educational Background

As shown in Table 3 most of the analysed job ads require a higher education degree (97%). A postgraduate degree such as PhD is required by 4% of the job ads. We observe that the number of job ads that prefer applicants specifically with a UAS degree compared to a GU degree is about the same (136 vs 144). Some job ads (61) do not distinguish between UAS and GU.

Table 3: Required Education level

Education	# Job Ads
Not requiring Higher Education	11
Higher Education - UAS	136
Higher Education (UAS or GU)	61
Higher Education - GU	144
Postgraduate Education (e.g. PhD)	15

Furthermore, we investigated which type of educational programs is preferred most. From the 125 job ads that we coded, 15 contained one or more preferred educational programs. AI, Mathematics and Computer Science are mentioned most: 12, 11 and 11 times respectively. Econometrics is mentioned twice, while physics and psychology—once. The report from LinkedIn [19] also confirms that the educational background of AI talent is quite diverse, but mostly related to IT: computer science, electrical and electronics engineering, computer and information science, and computer engineering.

4.4 Job Titles

We cleaned job titles by removing additional information in the title like a relation to the required seniority level or domain and by replacing full names with abbreviations, e.g. job titles like "Junior Machine Learning Engineer" or "AI Engineer Autonomous Vehicles" are counted with the job title "ML Engineer" and "AI Engineer", respectively. This resulted in 18 different job titles from the 367 selected AI engineer job ads, see Table 4. The most found job titles are "ML Engineer" and "AI engineer", where ML is used much more often than AI. "DL engineer" is mentioned 19 times and much less often than AI or ML engineer. Next to "AI/ML/DL engineer", the job titles "data scientist" and "data engineer" are frequently mentioned (respectively 23 and 19 times), and can therefore be regarded as jobs that in certain situations require AI engineering. Further note that some job titles even go into more detail by defining specific subareas of DL such as "computer vision" or "NLP (Natural Language Processing)".

We also remark that one of the the job titles in Table 4 is "Software Engineer". As opposed to the software engineering job ads that have been excluded (cf. Table 1), the job descriptions in the 18 software engineer ads included in Table 4 do refer to AI Engineering tasks or responsibilities.

Table 4 shows that a diversity of job titles is used for AI engineering jobs. Although according to our definition AI engineering encompasses both machine learning and software engineering, some job titles focus on one of these two. This makes it of interest to do a more detailed analysis of the job tasks in the job ads.

5 JOB CHARACTERISTICS OF AN AI ENGINEER

In this section we discuss our findings related to each of the four research questions.

Table 4: Distribution of categories of job titles

Position	# Job Ads
ML Engineer	173
AI Engineer	77
Data Scientist	26
Data Engineer	23
DL Engineer	19
Software Engineer	18
Computer Vision Engineer	9
NLP Engineer	3
Data Analyst	2
ML Scientist	2
ML Specialist	2
AI Consultant	2
AI Scientist	1
AI Specialist	1
AI Analyst Data Consultant	1
IT Engineer	1
R&D Engineer	1

5.1 RQ1a: Which Job Tasks Are Expected from an AI Engineer?

Using the two step coding approach, we found five main categories of tasks and responsibilities directly related to AI engineering: software development, modeling, data engineering, operations and business understanding.

Software development tasks and responsibilities. While advertised as ML engineering jobs, many phrases (151) are not directly related to ML techniques. Rather these phrases suggest tasks and responsibilities related to traditional software development: *e.g.,* "rewriting the research algorithms developed by your colleagues into efficient and stable Python code and scripts, wrap all into stable end-to-end solutions, for example Docker containers" and "develop these solutions, together with software developers, into intuitive and easy-to-use applications."

Modeling tasks and responsibilities. Another large group of phrases (100) is related to modeling tasks and responsibilities, *e.g.,* "extract insights from data and build ML models" and "design of deep neural networks." This category also includes tasks related to statistical modeling, *e.g.,* "use statistical models and algorithms."

Business understanding tasks and responsibilities. 57 phrases express tasks connected to the business needs and goals of the company, *e.g.,* "exploring and identifying opportunities, experimenting", "recommend ideas to business" and "working with business to prepare requirements."

Data engineering tasks and responsibilities. 44 phrases referred to data engineering tasks, *e.g.,* "building a data platform" and "translate business cases of our clients into usable, actionable data sets".

Operations-related tasks and responsibilities. Finally, the smallest group of phrases (39) referred to operations-related tasks and responsibilities such as being "able to bring these models into a production environment" and "conducts quality control of all models and systems to ensure an overall robust infrastructure".

Table 5: Focus of job tasks in job ads

Focus	# Job Ads
Data Science	35
Software Engineering	38
Data Science & Software Engineering	47
No info	5

5.2 RQ1b: Are Job Tasks Expected from AI Engineers More Focused on Data Science or Software Engineering?

Using the two step coding approach we analyzed the job ads for focus on data science or software engineering. Table 5 shows the results of this analysis. About forty percent of the job ads did not have a specific focus on either data science or engineering tasks, but instead required a mix of data science and engineering tasks.

Based on Table 5, we propose to distinguish three different profiles of AI engineers:

(1) Data Science Engineer: focus on data science job tasks;
(2) AI Software Engineer: focus on engineering job tasks;
(3) Generalist AI Engineer: both data science and engineering job tasks

Table 5 shows that about 40% of the investigated job ads asked for a generalist AI Engineer, and that most job ads are looking for specialist AI engineers, focusing on modeling or development tasks. This concurs with the observation of Lewis et al. [18] that three different roles (data scientist, software engineer and operations staff) often work together for building a production-ready machine learning system.

5.3 RQ2: What Are the Top Technologies Expected to Be Used by AI Engineers?

In this section we present or findings on programming languages, software technologies and AI engineering tools.

Software Programming Languages. Python was mentioned most often (36 times), followed by Java (11) and C++ (2). These findings match the LinkedIn report [19] that also found Python to be the most requested programming language, before C++ and Java.

Software Technologies. ML engineers are expected to master a broad spectrum of technologies ranging from software development to machine learning, and from cloud-related to data-related. Among the software technologies mentioned in the 125 job ads, the most popular ones are related to version control (Git has been mentioned 18 times), containerization (Docker/Kubernetes—10 times, and ECS/EKS—13 times), and JavaScript frameworks (Ember—11 times).

Among the cloud technologies, Azure is the most popular (9 job ads) followed by Amazon Web Services (6) and Google cloud (5).

Table 6: Soft skills

Soft skill	# Job Ads
Team-oriented	59
Open to learn	57
Coaching	32
Passionate	15
Result-driven	12
Analytical	11
Innovative	10
Communicative	9
Creative	5
Curious	4

Tensorflow is the most commonly mentioned ML technology (6 times), with SKLean including SciKit (5), Keras (3), Pandas (3) and Torch (3) completing the top 5. These AI and data-science libraries are also among the top ones found by the LinkedIn report [19].

Finally, the most mentioned data technologies are Spark (12 times) and SQL (7).

AI Engineering Tools. Such websites as *Awesome Production Machine Learning*[2] identify several tools that support standard software engineering activities like version control (DVC), containerization (KubeFlow) or configuration management (MLFlow, Comet) for AI-enabled systems. These tools are hardly mentioned in the job ads we investigated (MLFlow is the only exception and it is only mentioned three times in all 367 job ads). As specific AI engineering tools such as DVC, MLFlow and KubeFlow are relatively new, it is not surprising that we did not find them in the job descriptions yet.

5.4 RQ3: Which Soft Skills Are Required for the Job of AI Engineer?

In the 125 job ads that we analyzed with respect to soft skills, there is no separate section about required soft skills. Mostly, soft skills are mentioned somewhere in the function description, e.g., "Clearly communicate actionable information to the project team". From the 125 coded job ads on soft skills, 59 explicitly mention team work. Also willingness and open to learn is mentioned often, with 57 times. Coaching and mentoring are mentioned in 32 job ads. Other soft skills are mentioned less often, see Table 6.

The set of soft skills we found in the job ads concurs with the set of soft skills found for software engineers by Matturro et al. [21] or for requirements engineers by Daneva et al. [8]. In our job ads we find a high frequency of "open to learn", while Matturro et al. [21] found "willingness to learn" to score at the low side in their systematic mapping study of soft skills considered relevant in software engineering. This might be due to the fact that machine learning (or AI in general) is such a new and fast-moving field that companies put emphasis on this learning ability.

[2] https://github.com/EthicalML/awesome-production-machine-learning

6 DISCUSSION

In this section we discuss implications of our work for AI engineering research, AI engineering education, and for AI Engineering in industry.

6.1 Implications for AI Engineering Research

In Section 5.1 we presented five main categories of tasks and responsibilities for the AI engineer: software development, modeling, data engineering, operations and business understanding. The first four of these five categories match the phases as depicted in Figure 1, but business understanding is not a separate phase in the life-cycle as proposed by Farah [9] and Lwakatare et al. [20].

The category name "business understanding" stems from the CRISP-DM data science process [6]. However, this process does not include software development activities, where we found a software development job tasks category in the job ads. Case studies by Amershi et al. [1] and John et al. [13] also present ML development processes that include requirements engineering steps. As Amershi et al. [1] state: "In the model requirements stage, designers decide which features are feasible to implement with machine learning and which can be useful for a given existing product or for a new one. Most importantly, in this stage, they also decide what types of models are most appropriate for the given problem". Both Amershi et al. and John et al. do not discuss software requirements engineering or other software development activities in their ML development process.

If we look at the categories of job tasks we found in the job ads (see Section 5.1), the AI engineering life-cycle as proposed by Farah [9] and Lwakatare et al. [20] most closely matches, because it does include software development steps. However, we propose to extend it by adding a "Business Understanding (BUS)" phase as depicted in Figure 2.

AI engineering research should continue to work on defining AI engineering and the AI engineering life-cycle with AI engineering job tasks. In the job ads we see the need for both data science engineers and AI software engineers, indicating that both disciplines (data science and software engineering) are needed in AI engineering research.

6.2 Implications for Educating AI Engineers

In the job ads we saw that AI engineers from both Universities of Applied Science and General Universities are almost equally sought for. This means that both application-oriented and scientifically-oriented universities should offer study programs for AI engineering, on both bachelor and master level. Both the LinkedIn report [19] and our analysis indicate that these AI engineering programs are most likely to be offered by educational institutions that currently offer IT-related studies such as computer science, electrical and electronics engineering, computer and information science, and computer engineering.

The content of these study programs should be aligned with the different AI engineering job profiles we found in Section 5.2. An AI engineering program could focus on one of the two specialist profiles (data science or software engineering), but there should also be programs that cater for the generalist AI engineer profile. For example, at Fontys UAS (the institution of the first and the second

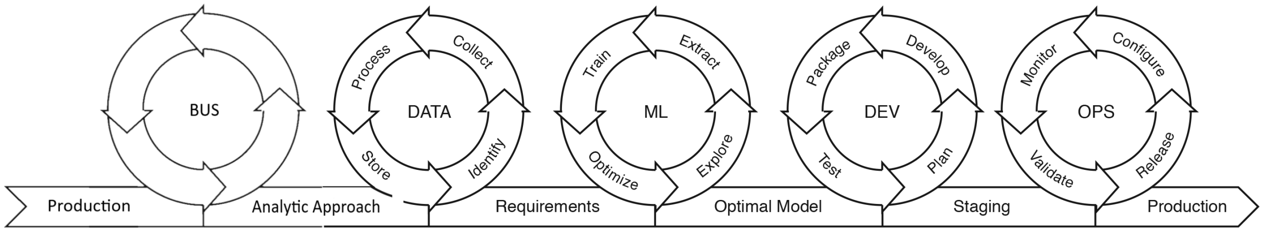

Figure 2: The extended AI Engineering life cycle, based on Farah [9], with a Business Understanding (BUS) phase

author) we currently offer an AI specialization (2 semesters) that together with the software engineering basic profile (6 semesters) educates a bachelor-level applied AI Software Engineer (4 years). This is a specialist AI engineer education focusing on software engineering, which shows this type of students can work on AI engineering without deep mathematical background, but with a basic understanding of machine learning and its applications [11, 12].

Future work remains to further investigate the three AI engineer profiles: are there differences in soft skills demanded or technologies? Is a different educational background required? With the current job ad set we have, we cannot answer such questions satisfyingly.

The extended AI engineering life-cycle we present in Figure 2 could help educate AI engineers that understand the complete AI engineering life-cycle, even though the education might focus on specific phases of the life-cycle. The life-cycle shows that to build a production-ready machine learning system the AI engineer needs to be able to work with data, models and software.

With respect to soft skills we conclude that the AI engineering programs could make use of the existing soft skills education for engineers, as we did not find any new soft skills in the job ads. From our data, learning ability seems an important soft skill for AI engineers. The fact that the discipline is still evolving, makes this mastering of new insights (tools, publications, models, etc.) part of the education by default.

According to our study of the job ads, the AI engineering programs should select a combination of software engineering (e.g. Python, Git, Kubernetes, Cloud, SQL) and machine learning (e.g. Tensorflow, scikit-learn, MLFlow) tooling and technology. Note that AI engineering tool support is an area where rapid developments are to be expected [17].

The job ads we analyzed indicate that a diversity of domains is looking for AI engineers, see Table 2. Many job ads we saw contain job tasks in the category "Business Understanding": working with domain experts to create valuable AI solutions for the company. To facilitate this cooperation we not only need to train AI engineers on requirements engineering but also domain experts on the possibilities of AI for their domain. Thus, there is a need to offer modules or even separate programs on the application of AI-enabled systems in a given domain.

6.3 Implications for Industry

Data scientists and software engineers looking for a job in AI engineering and companies advertising jobs in AI engineering should be aware of the three different AI engineer profiles that we identified in Section 5.2. Furthermore they should be aware of the diversity of job titles use in this discipline. For both employer and employee it is thus important to look at the job tasks in detail. For this they could use the extended AI engineering life-cycle we present in Figure 2. This Figure indicates the diversity of job tasks that belong to an AI engineering project and could help to discuss the focus of job tasks for a given AI engineer position.

As said in Section 5.2 most job ads we saw are for specialist AI engineers that focus on modeling or software development. They would thus need to work in a team with domain experts, data scientists, software engineers, data engineers or operations engineers to execute all AI engineering job tasks. The job task categories we identified in Section 5.1 and the extended life-cycle we present in Figure 2 could help to form AI engineering teams in such a way all tasks are covered by the team members.

Our research indicates a growing demand for AI engineers in the coming years. These AI engineers need to be educated. At the same time AI engineering is a new and evolving discipline. Our question "What do companies require from AI Engineers?" remains valid for the coming years, as education will need to be updated when the discipline evolves. Industry should work together with universities in both the data science and software engineering disciplines to ensure education meets these changing demands.

7 THREATS TO VALIDITY

As with any empirical study the validity of our conclusions might be threatened. In the following sections, we describe threats to validity pertaining to dataset selection, job ad selection and job ad coding. We also discuss what we did to mitigate the risks that these threats pose.

7.1 Job Ad Selection

The study has focused on job ads from The Netherlands, and hence generalizability of the findings might be a concern. However, The Netherlands is strongly oriented towards the international labor market, which can be seen from the fact that about one third of the job ads is written in English. The related reports by Indeed (USA) [22] and LinkedIn (Europe) [19] indeed show similar results for other regions. We took great care to confirm all of our findings from other sources, avoiding bias for the Dutch situation. This is why we expect our conclusions to hold for other countries. However, future replication studies should verify this expectation.

In order to be as objective as possible about which job ads to select, we followed the guidelines of Kitchenham and Charters [16] to define selection criteria and a search query on beforehand.

The final selection of 367 job ads was performed by the first author and checked by the second author, as explained in Section 3. The first author holds a PDEng in Software Engineering and the second author a PhD. Both authors have experience in AI engineering through research, educational tasks and graduation supervision. The discussion between the two authors resulted in a common understanding why a job ad should be included or excluded.

We filtered out exact duplicates in the job ads, but still some of the companies in the investigated set of 367 job ads occur more than once. This could bias the result. However when we look at the top three of duplicate companies, we see 6 job ads from financial company X, 5 job ads from consultancy company Y and 4 job ads from high-tech company Z1, 4 job ads from food company Z2 and 4 job ads from startup Z3. This indicates that bias towards a certain company or industry/sector based on duplicate job ads is not unreasonably large.

7.2 Job Ad Coding

The manual coding for job tasks, technologies, soft skills and educational background in the job ads was performed by the first author and checked by the second author. When the second author felt a certain phrase was missed or coded in the wrong category, a rule was defined on the basis of which all job ads and found phrases where re-coded. This process was repeated until no further disagreements where found between first and second author.

Another countermeasure we took for this subjective step is to ensure our category names matched with existing terminology from Farah [9] and CRISP-DM [6]: business understanding, data (engineering), modeling, (software) development, operations.

There is a possibility that we missed some phrases in the job ad descriptions that thus have not been coded at all. To mitigate this we coded several job ads with all three authors until we agreed on the way of coding the job ads. After this we were quite confident that the first author would be complete enough in the coding process.

8 CONCLUSION

In this paper we define AI engineering as a combination of software engineering and machine learning to build production-ready machine learning systems. We have analysed 367 job ads to learn more about the profession of AI engineer. This paper contributes to the understanding of AI engineering in the following ways:

- We confirm that within AI engineering there is a diversity of job titles that go beyond AI, ML or DL engineer.
- We identify five job task categories for the AI engineer: software development, modeling, data engineering, operations and business understanding.
- We present an extended AI engineering life cycle with a business understanding phase.
- We note that there is not a wide-spread use yet of specific AI engineering tools, instead we see a combination of software engineering tools and AI tools.
- We confirm learning ability as an indispensable soft skill for AI engineers.
- We identify the need for AI engineers from both UAS and GU.
- We define three different profiles for the AI engineer: 1) data science engineer with focus on modeling, 2) AI software engineer with focus on software development, 3) generalist AI engineer with focus on both modeling and software development.
- We present implications for AI Engineering research, education and industry.

As we also saw during our study, the number of AI engineering related vacancies will continue to grow throughout the coming years. It is a huge challenge to educate these AI engineers. In our opinion both data science and software engineering programs should educate professionals that are knowledgeable on the entire AI engineering life-cycle. These professionals should form multidisciplinary teams to build production-ready machine learning systems. Our future work lies in further investigating the different profiles of AI engineers and their education, and in investigating development processes and supporting tools for AI engineering teams.

ACKNOWLEDGMENTS

We would like to thank Gerard Schouten and Fontys University of Applied Sciences for supporting our research.

REFERENCES

[1] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 291–300.

[2] Jan Bosch, Ivica Crnkovic, and Helena Holmström Olsson. 2020. Engineering AI Systems: A Research Agenda. *arXiv* (2020), 1–8. arXiv:2001.07522

[3] Andriy Burkov. 2020. *Machine Learning Engineering*. True Positive Incorporated.

[4] Anita D Carleton, Erin Harper, Tim Menzies, Tao Xie, Sigrid Eldh, and Michael R Lyu. 2020. The AI Effect: Working at the Intersection of AI and SE. *IEEE Software* 37, 4 (2020), 26–35.

[5] Maura Cerioli, Maurizio Leotta, and Filippo Ricca. 2020. What 5 Million Job Advertisements Tell Us about Testing : a Preliminary Empirical Investigation. In *35th ACM/SIGAPPSymposium Appl. Comput. (SAC '20), March 30-April 3, 2020, Brno, Czech Repub.* ACM, New York, NY, USA, 1586–1594. https://doi.org/10.1145/3341105.3373961Permission

[6] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, and R. Wirth. 2000. CRISP-DM 1.0 Step-by-step data mining guides.

[7] Lito Perez Cruz. 2017. When Data Science Becomes Software Engineering. In *In Proceedings of the 9th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management (KEOD)*. 226–232.

[8] Maya Daneva, Chong Wang, and Patrick Hoener. 2017. What the Job Market Wants from Requirements Engineers? An Empirical Analysis of Online Job Ads from the Netherlands. *Int. Symp. Empir. Softw. Eng. Meas.* 2017-Novem (2017), 448–453. https://doi.org/10.1109/ESEM.2017.60

[9] Danny Farah. 2020. The Modern MLOps Blueprint. https://medium.com/slalom-data-analytics/the-modern-mlops-blueprint-c8322af69d21.

[10] Google. [n.d.]. MLOps: Continuous delivery and automation pipelines in machine learning. https://cloud.google.com/architecture/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning.

[11] Petra Heck and Gerard Schouten. 2020. Turning software engineers into AI engineers. *arXiv preprint arXiv:2011.01590* (2020).

[12] Petra Heck and Gerard Schouten. 2021. Lessons learned from educating AI engineers. In *2021 IEEE/ACM 1st Workshop on AI Engineering-Software Engineering for AI (WAIN)*. IEEE, 1–4.

[13] Meenu Mary John, Helena Holmström Olsson, and Jan Bosch. 2020. Developing ML/DL Models: A Design Framework. In *Proceedings of the International Conference on Software and System Processes* (Seoul, Republic of Korea) *(ICSSP '20)*. Association for Computing Machinery, New York, NY, USA, 1–10. https://doi.org/10.1145/3379177.3388892

[14] Mary Anne Kennan, Fletcher Cole, Patricia Willard, Concepción Wilson, and Linda Marion. 2006. Changing workplace demands: What job ads tell us. *Aslib Proc. New Inf. Perspect.* 58, 3 (2006), 179–196. https://doi.org/10.1108/00012530610677228

[15] Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2018. Data Scientists in Software Teams: State of the Art and Challenges. *IEEE Transactions on Software Engineering* 44, 11 (2018), 1024–1038. https://doi.org/10.1109/TSE.2017.2754374

[16] B. Kitchenham and S Charters. 2007. *Guidelines for performing systematic literature reviews in software engineering*. Technical Report. EBSE-2007-01.

[17] Peter Krensky, Carlie Idoine, Erick Brethenoux, Pieter Den Hamer, Farhan Choudhary, Afraz Jaffri, and Shubhangi Vashisth. 2021. Magic Quadrant for data science and machine-learning platforms. *Gartner, Inc* (2021).

[18] Grace A Lewis, Stephany Bellomo, and Ipek Ozkaya. 2021. Characterizing and Detecting Mismatch in Machine-Learning-Enabled Systems. *WAIN21* (2021).

[19] LinkedIn. 2019. AI Talent in the European Labour Market. https://economicgraph.linkedin.com/content/dam/me/economicgraph/en-us/reference-cards/research/2019/LinkedIn-AI-Talent-in-the-European-Labour-Market.pdf.

[20] Lucy Ellen Lwakatare, Ivica Crnkovic, and Jan Bosch. 2020. DevOps for AI–Challenges in Development of AI-enabled Applications. In *2020 International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. IEEE, 1–6.

[21] Gerardo Matturro, Florencia Raschetti, and Carina Fontán. 2019. A systematic mapping study on soft skills in software engineering. *JUCS-Journal of Universal Computer Science* 25 (2019), 16.

[22] Sarah Overmyer. 2018. Jobs of the Future: Emerging Trends in Artificial Intelligence. https://www.indeed.com/lead/artificial-intelligence-report.

[23] Maria Papoutsoglou, Nikolaos Mittas, and Lefteris Angelis. 2017. Mining people analytics from stackoverflow job advertisements. *Proc. - 43rd Euromicro Conf. Softw. Eng. Adv. Appl. SEAA 2017* (2017), 108–115. https://doi.org/10.1109/SEAA.2017.50

[24] Anat Rafaeli and Amalya L. Oliver. 1998. Employment ads: A configurational research agenda. *J. Manag. Inq.* 7, 4 (1998), 342–358. https://doi.org/10.1177/105649269874009

[25] Johnny Saldaña. 2011. *The Coding Manual for Qualitative Researchers* (2nd editio ed.). SAGE Publications Inc. 329 pages.

[26] Danilo Sato, Arif Wider, and Christoph Windheuser. 2019. Continuous delivery for machine learning. https://martinfowler.com/articles/cd4ml.html.

[27] Sami Surakka. 2005. Analysis of Technical Skills in Job Advertisements Targeted at Software Developers. *Informatics Educ.* 4, 1 (2005), 102–122. https://doi.org/10.15388/infedu.2005.07