Labor market analysis and curriculum gap assessment using big data in Kenya

Final report for the World Bank contract 7192067

Harri Ketamo and Anu Passi-Rauste 27.9.2019

Headai Ltd. Finland
# Table of Contents

- Executive summary  4
- Introduction  6
  - Context  6
  - Objectives  7
  - Project implementation in a nutshell  7
  - Headai and the team  9
- Methodology  10
  - Digital Twin  10
  - Data sources  11
    - Job advertisement data  11
    - Curriculum and educational offering in Kenyan universities  12
- Outcomes, results, and findings  13
  - Kenyan labor market skills demand  13
    - Key insights - TOP 10 Industries, TOP 20 jobs and skills  14
  - Kenyan software and ICT industry skills demand  17
    - Similarities and differences between skills demand and educational offering  21
      - Moi University  21
      - Feedback from MOI University  25
      - University of Nairobi  26
- Summary and recommendations for the future  30
  - Suggestions for further opportunities  31
- References  33
List of figures

Figure 1: The skills demand data visualized as a skills map ............................................................ 7
Figure 2: The curriculum offering data visualized as a skills map .................................................. 8
Figure 3: The gap analysis visualized as a skills map ...................................................................... 8
Figure 4: Microcompetencies is a platform and service to use AI-based methods for skills identification and labor market analysis. https://www.microcompetencies.com/api .................. 9
Figure 5: The visualization of teh overall skills demand in Kenya ................................................ 13
Figure 6: The skills demand in Kenya in the following domains: SOFTWARE, ICT, PROGRAMMING, JAVA, PYTHON, SCRIPT, JAVASCRIPT, SQL, DATA, NETWORK, SERVER, BACKEND, FRONTEND, CLOUD COMPUTING, CLOUD, HTML, AGILE ........................................ 17
Figure 7: Skills taught at the University of Nairobi’s computer science program ........................ 19
Figure 8: Skills taught at Moi University in software and ICT from 2018-2019 ............................ 20
Figure 9: Technical skills compared to business skills taught at Moi Unievristy ............................. 20
Figure 10: Technical skills cmopared to business skills taught at University of Nairobi ............. 21
Figure 11: Skills taught at Moi university in software and ICT. Similarities and differences between skills demand and educational offering ................................................................. 21
Figure 12: Skills taught at Moi University in software & ICT: Skills gaps category 1: Professional Software Developer .......................................................... 22
Figure 13: Skills taught at Moi University in software & ICT: Skills gaps category 2: Industry platforms & tools .......................................................... 23
Figure 14: Skills taught at Moi University in software & ICT: Skills gaps category 3: Financing & business-related ICT skills .......................................................... 24
Figure 15: Skills taught at Moi University in software & ICT: Skills gaps category 4: Basic ICT skills .......................................................... 25
Figure 16: Skills taught at University of Nairobi in software and ICT. Similarities & differences between skills demand and educational offering ................................................................. 26
Figure 17: Skills taught at University of Nairobi in software and ICT: Skills gap category 1: Professional Software Developer .......................................................... 27
Figure 18: Skills taught at University of Nairobi in software and ICT: Skills gaps category 2: Industry platforms & tools .......................................................... 28
Figure 19: Skills taught at University of Nairobi in software and ICT: Skills gaps category 3: Financing & business-related ICT skills .......................................................... 29
Figure 20: Skills taught at University of Nairobi in software and ICT: Skills gaps category 4: Basic ICT skills .......................................................... 30

List of tables

Table 1: Top 10 industries during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017) .......................................................... 14
Table 2: Top 20 job titles during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017) .......................................................... 15
Table 3: Job title relations to the title Provider ............................................................................. 16
Table 4: Top 20 skills during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017) .......................................................... 16
Table 5: TOP 20 skills in Kenyan software & ICT industry skills demand ..................................... 18
List of acronyms

AI: Artificial Intelligence
API: Application program interface
ESCO: European Skills, Competences, Qualifications and Occupations
ICT: Information and Communications Technology
IT: Information Technology
NLP: Natural Language Processing
SOM: Self-Organizing Map
Executive summary

One of the biggest challenges in designing a labour market relevant curriculum is to make sure that the curriculum is up to date and uses the same vocabulary as the employers do. This has been relatively a little studied challenge: if we take almost any curriculum, it is written in an academic language. At the same time, curriculum and course descriptions are lacking a description of skills that students can gain and present when seeking for a job, especially through online job portals.

In this study, Cognitive AI, Big Data and Natural Language Processing were brought together to build a real-time understanding on skills, competencies, knowledge and abilities that employers are looking for. This information is visualized as maps that enable us to: i) understand what skills are needed now and in the near future; ii) guide up-to-date curriculum development; iii) advise students on their course selection; and iv) gain an understanding to improve the competitive offering of universities. From methodological point of view, this is also a pilot study how we use modern technologies, like cognitive AI and Natural Language Processing (NLP) in predictive analysis, and what factors we should take into account when designing future studies with these technologies.

This study used Kenya as an example and the result shows the similar trend in Kenyan universities as observed among European universities: Kenyan universities do have good academic curriculums, but there are several ways to enhance the curriculum and course descriptions that can better address the needs of the labour market. In some cases, adding the terminology into course description would be enough. This study also analysed the top industries, skills and jobs presented in the online job posting data. From the data, we could show historical trends and current skill needs as well as predict what skills or jobs are down-trending, i.e. what skills and jobs are most likely not to be sought in the near future. In addition, the findings showed that most of the job advertisement was related to digital and Information and Communications Technology (ICT) skills. The most trending jobs in the past few years included officer, manager, and assistant.

A significance of the study is that new AI-based technologies can analyze the labor market demand quickly and help universities review their curriculum and name their courses to match with the vocabulary, semantics, terminologies and language used in the labor market. This enables university students to identify the skills they learn at universities and present them effectively to match the skills demanded in the labor market.

While the technology contributed to making the labor market analysis faster and easier, the challenges still remain. Due to a technical nature of the AI-based tool, results of the analysis are not easily understood by non-technical people. Thus, more user-friendly interface may be needed to better digest the information and communicate the results effectively.

We are sincerely grateful to the World Bank and especially to Saori Imaizumi for her professional commitment and a strong interest in the topic as well as her inputs for this report and the design of the project. We also appreciate Ruth Charo and Moi university for their participation in the virtual meeting and feedback to the study. We look forward to sharing and
discussing the study and results with a broader audience at The World Bank as well as with other stakeholders.

We have only taken the first steps of applying AI to educational development. In this study, we have applied Headai technology and a “digital twin” concept to develop a digital replica of the skills demand and supply in a form of a skills map. Digital twin has been used in automotive, construction, manufacturing, utilities, and healthcare industries to create a digital or virtual copy of physical assets and products. Digital twin connects between virtual and physical world with real time data collected through sensors. Using this data, simulation has been developed. Headai applied this concept into labor market and skills analysis. Digital twin also expands as a development concept in many industries. We would like to see education industry adapting it and enabling us to move towards transparent predictive analytics, cognitive reasoning in any language.

**Keywords:** Data-driven labor market analysis, AI-based skills maps, Curriculum development, Artificial Intelligence, Big data analytics, Skills mismatch, Digital Twin
Introduction

Context

Our society is facing the challenges of continuous change, global competition, digitalization, and the replacement of human labor by smart automation. At the same time, companies are reporting a lack of skilled workers and global skills mismatches. This does not mean that there are no skills on offer – rather, available skills do not have enough demand. In addition, skills have not been defined in a commensurable manner between different actors, not even in terms of the terminology used, not to mention the precise semantic meanings of these terms.

Company’s growth depends on skills. They need the resources and skills of strong partners in addition to their own competitive advantage. Growth arises in ecosystems when the right parties find each other. We want to ensure that there is enough talent for businesses, and companies will have access to the required skills to grow and develop into the best in the world.

Technological transformation has already had a major impact on the demand for skills in the labour market. The skills required to enter and progress in the labour market are undergoing profound changes. It is not only the occupational structure that is changing – the changes apply to skills requirements within occupations as well. Education, training and learning are the most critical means of coping with the transformation of work. Continuous learning, upskilling and reskilling are becoming more and more important.

One of the biggest challenges for the education institutes in the design of a labor market related curriculum is to make sure that the curriculum is up-to-date and uses the same vocabulary as the labor market. This may sound easy, but almost any curriculum is written in an academic language and, at the same time, may lack terms that are currently used in the labor market, which may prevent students from appropriately presenting the skills they learn at schools in the job market.

Every target group; companies, industries, and educational institutions, are facing the challenges of speed and adaptability of constant change in the world – how can they structure the complex, occasionally chaotic, operating environment into a comprehensible form? The era of the straightforward and predictable operating environment is over for good. Labor markets as well as education providers are seeking solutions to enable them to better identify and utilize skills. Skills cannot be utilized if they are not first identified. Nor can they be identified without a common language and terminology.

“Without a common language there can be no discussion” Wittgenstein

The recruitment of new employees is happening more and more via digital channels, such as social media, magazines, newspapers, requirement sites and companies’ own websites. Companies also produce blog articles, white papers and social media posts where they indicate their latest skills needs while presenting new products, technologies, services and customer testimonials. Those textual material delivered and shared via digital channels creates huge open data sets that can be used by schools and curriculum designers. It fulfils the typical
requirements of big data. Universities and education institutes can gain similar value in analyzing open data from the future employers they serve.

Advanced analytics and artificial intelligence-based tools and technologies provide new ways to discover skills demand and provide valuable information for different stakeholders. These tools have been tested in middle-income and high-income countries, but not in Sub-Saharan Africa. As the data sources are expanding also in Sub-Saharan Africa, this project aimed to study the potential for leveraging the AI-tools to provide relevant skills demand information for students and job seekers to better train themselves to increase the chance of employment. Moreover, policy makers, universities and training institutions can also use the information to offer market relevant curriculum and courses.

**Objectives**

The objective of this work was to leverage AI tool to conduct a labor market demand analysis and curriculum gap assessment for two universities in Kenya. This is a pilot study and focuses especially on the software and ICT industry which tends to have more uniform occupational standards. It provides knowledge on how AI and data analytics can be used and provide value for the curriculum (re)design at the university to meet the requirements of employers and different industries as well as develop new short courses to fill the labor market demand and supply gap.

**Project implementation in a nutshell**

The project was conducted during few intensive months (May-July 2019). It included following phases:

1. **Headai** collected and analyzed online job advertisement data using Headai AI-enabled tool from three selected local, public and openly machine-readable online job boards available in Kenya.
   - [Jobwebkenya](https://jobwebkenya.com/)
   - [Kenyancareer](http://www.kenyancareer.com)
   - [Jobskenyaone](https://www.jobskenyaone.com/)

2. **Headai** AI-enabled tool visualized and created skills map, which focused on digital and ICT skills.

![Figure 1: The skills demand data visualized as a skills map](image)
3. Headai machine-read and analyzed two university curriculums;
   ○ University of Nairobi, Degrees offered in School of Computing and Informatics
   ○ Moi University, Degrees offered in ICT Department

4. Headai conducted a gap analysis between the labor market and the curriculums to assess the digital skills gap.

5. Headai shared the analyzed data for future use in a presentation format.

6. The World Bank and Headai together conducted a virtual workshop to share the results with Moi University. (Unfortunately the virtual workshop with University of Nairobi could not be materialized.)

7. Headai compiled the results into this report combined with a visual presentation.
**Headai and the team**

The World Bank chose Headai to conduct the study. Headai [http://headai.com/](http://headai.com/) is a technology company from Finland with a mission to scale goal-oriented professional development for everyone.

**Headai tech stack**

Headai’s cognitive artificial intelligence, AI, is 100% owned by the company and in commercial use internationally since 2015. It is based on self-organizing semantic network which enables more complex reasoning (with Natural Language Processing) than traditional ontologies (eg. O*net) and methods. It structures complex job market related data into a (visually) comprehensible form with its unique AI system.

The AI service can also be utilized via the REST API (an application program interface). APIs enable a standard mechanism to share data and functionality (see Figure 4). This is a user-friendly, easy and fast way of using Headai AI methods and does not require great programming skills.

**Headai team**

Headai team for the assignment consisted of Harri Ketamo as a Chief Data Scientist and Anu Passi-Rauste as a Project Director. Both team members have extensive background in the main disciplines of the assignment.

Harri Ketamo (PhD), [https://www.linkedin.com/in/harriketamo/](https://www.linkedin.com/in/harriketamo/) founder and chairman at Headai. He is a Data Science lead in Headai AI solutions and customers data analyses. He has been working in the field of EdTech and games since 90’s. He published +100 scientific papers and presented +250 talks on AI in predictive workforce and professional development. Harri Ketamo was responsible for data analysis from online job portals and curriculums, AI development, and sharing the results for the customer.

Anu Passi-Rauste, [https://fi.linkedin.com/in/anupassirauste](https://fi.linkedin.com/in/anupassirauste), has 20+ experience in Edtech, Human Resource Development, Workforce Development, Education, and International Development projects (EU, Ministry for Foreign Affairs of Finland). Her key role was in developing and validating Headai’s AI solutions with user-customers globally. Anu Passi-Rauste was responsible for coordinating and delivering the assignment and results as planned with the tight schedule. Anu was also responsible for communicating with customers, both World Bank and identified organisations in Kenya.

![Figure 4: Microcompetencies is a platform and service to use AI-based methods for skills identification and labor market analysis.](https://www.microcompetencies.com/api)
Methodology

This pilot study used Headai’s cognitive artificial intelligence, that deploys Natural Language Processing (NLP) algorithms, to build and develop text based digital twins (digital replica) on knowledge domains, groups, and persons. This enables transparent predictive analytics, cognitive reasoning in any language, in any domain.

Digital Twin

Headai uses the concept of Digital_Twin in order to build its dynamic knowledge graph about language. Digital Twin is commonly used concept in domains like Industry 4.0, Internet of things and 5G; “originally developed to improve manufacturing processes, digital twins are being redefined as digital replications of living as well as non-living entities that enable data to be seamlessly transmitted between the physical and virtual worlds.”

In short, in Headai’s methodology Digital_Twin is a digital replica on selected domain’s (language based) understanding / knowledge.

Headai’s Digital_Twin is constructed by applying unsupervised (a Self-Organizing Map (SOM), SOM type) machine learning, which enables us to build always up-to-date and detailed language models in different contexts. Headai’s autonomous-learning artificial intelligence builds models itself and uses them to explain the world and make decisions (known as the third wave of artificial intelligence: https://lounge.fim.com/tekoaly/).

This language model is applied in this project to build a detailed and extremely flexible domain specific Digital_Twins on job demand and curriculum. Headai also builds Digital_Twin on learner/worker/trainer.

In this project, we call visualized Digital_Twins also as a skills map, where the map topology presents in one image with the fuzzy relations between single skills, how labor market demand is clustered as well as the demand volumes.

These skills maps enable us to:

i. understand what skills are in demand now and in the near future and predict changes to skills demand in general
ii. guide the curriculum development and updating process
iii. gain an understanding to improve the competitive course offerings of the universities

Natural language processing is more than a keyword search. It is dealing with the semantics and ontology of the information. The work-related semantics consist of 1.5 million exact relations, more than 5 million general relations and more than 20 million loose relations among the skills, occupations and other relevant words. From an intellectual capital point of view, the value added is in extending the understanding to the dynamics of the language. For instance, most common Finnish manually-created ontologies consist of 20,000 – 80,000 words and 0.1 - 1 million relations. Still, it has taken tens of years and hundreds of work-years to build these ontologies.

1 https://ieeexplore.ieee.org/document/8424832
One way to call our AI's understanding of skills is also an 'ontology.' Basically, there is no big conceptual differences between ontology and knowledge graph. While ontology is usually a hand-made subjective set of concepts and relations, a knowledge graph is more detailed, and data driven view on the same phenomena.

Headai AI has been trained to recognise skills, competencies, attitudes, job titles and other relevant concepts from natural language and non-structured data. We have done that for several years and our AI is all the time up to date, because every time there is anything new (that AI does not know in advance), it will ask the trainer how the new concept should be understood. The Headai’s terminology and the relationships between terms – the ontology – connect micro skills and related validation to internationally recognized standards and ontologies, such as European Skills, Competences, Qualifications and Occupations (ESCO) or O*net. Both of these ontologies are subsets of our knowledge graph of the labor market.

Currently, Headai dynamic skills ontology has over 80,000 meaningful words and relations related to workforce and it is developing every day.

| Skills* | 10,108 |
| Relevance** | 23,782 |
| Relations | 615,533 |

*based on Headai AI training, compatible with ESCO
**words that do computationally explain something about the context

Data sources

The pilot study used two levels of data sets;

1. online job advertisement data from various sources, covering the skills demand in Kenya at a given time
2. curriculum and educational offering in Kenyan universities

The data are used for research and modelling purposes in the same manner as a human research team would use it: data are collected, cleaned, and classified; the data are analyzed; and the outcomes are reported. The only difference to manually-created research is the volume of data; we analyzed +60,000 unique public job openings. Artificial intelligence carried out the work of more than a hundred experts in almost real time at the cost of only a few dozen experts.

We also want to highlight that Headai AI uses the data as a researcher (human researcher) would do: read it through, take notes and records based on data and then leave the data as it is. A researcher would not distribute, share or modify the data nor the researcher would not own the original data. The same goes for our AI.

Job advertisement data

There are at least 20 online job portals in Kenya including some mobile apps. Headai conducted a research and looked at those different online job portals especially in Kenya (eg. Careerjet, Careers24, Labourmarket.go.ke, BrighterMonday, Fuzu, and UN Jobs). Global portals like LinkedIn also include Kenyan jobs but their job data is only open for selected partners.
To leverage AI tools for the work, we were looking for a large amount of online job advertisements (tens of thousands of jobs) from a local public website, which possess continuously updated directory that has traffic as well as openly machine-readable job advertisements. Based on these criteria, following labour market data sources were selected for the pilot.

Jobwebkenya https://jobwebkenya.com/
Kenyancareer http://www.kenyancareer.com/
Jobskenyaone https://www.jobskenyaone.com/

These job portals provide unstructured and natural language-based job openings. After data cleaning, there were over 60,000 job openings available for the analysis during the year 2015 and 2019, which enabled us to develop good and detailed data model about skills demand. If this text extraction had been done manually, it would have required more than 25 researchers working for a year. In each year, the following number of online job advertisements was available through the portal.

2019 => 21,500  
2018 => 23,000  
2017 => 11,000  
older => 7,500

Curriculum and educational offering in Kenyan universities

For the curriculum analysis, Headai analyzed Kenyan university curriculums. Only a few universities provide comprehensive curriculum and quality course descriptions available online. To enable a quick pilot study, Headai agreed with the World Bank to select universities that have quality data on their public website. Based on these criteria, we selected the University of Nairobi and Moi University for this pilot. The work focuses specifically on digital skills. Therefore, the School of Computer Science was selected as a target program. We also want to highlight that the University of Nairobi is the top university in Kenya and its computer science department is very well known. Also, this work aligns with the current initiative on university benchmarking, in which a participating university, Moi University, is involved.

Following curriculums were selected for the pilot (available in Internet in June 2019):

University of Nairobi, Degrees Offered in School of Computing and Informatics

- BSc. Computer Science²
- MSc. in Applied Computing³
- MSc. in Computational Intelligence⁴
- MSc. in Distributed Computing Technology⁵
- MSc. in Information Technology Management⁶

² https://sci.uonbi.ac.ke/index.php?q=uon_degrees_display/undergraduate  
³ https://sci.uonbi.ac.ke/index.php?q=uon_degrees_details/928  
⁴ https://sci.uonbi.ac.ke/index.php?q=uon_degrees_details/931  
⁵ https://sci.uonbi.ac.ke/index.php?q=uon_degrees_details/932  
⁶ https://sci.uonbi.ac.ke/index.php?q=uon_degrees_details/933
Moi University, Degree offered in ICT Department
- BSc. Information Sciences
- MSc. In Information Technology
(Only the curriculum overview and course structure, no detailed course descriptions available)

The data available from the curriculums was unstructured and natural language based. After cleaning the data, there were more than 100 relevant course items described for the model, which enabled good proof of concepts on the topic. However, the data is not enough to draw generalized conclusions on Kenyan higher education.

Outcomes, results, and findings

In general, if we do not know what skills are currently sought by employers, we cannot adjust the curriculum to meet the labor market and workplace needs. The purpose of the study was to show the gap between the curriculum and skills demanded in the labor market. The focus of this section is to show the possibilities of AI-generated data analysis and visualizations, not to test a specific hypothesis or present scientific proof.

Kenyan labor market skills demand

Figure 5: The visualization of the overall skills demand in Kenya
The overall skills demand in Kenya are visualized in the Figure 5. The basic unit is one skill, which is represented by a block. The number in each box explains the amount of relations that the skill word has with other words. The more a skill is sought, the deeper the green of the block becomes. The figure is clustered with an applied k-nearest clustering algorithm, i.e. the origin in the middle of the figure is the skill that has the most connections to other skills. The next connected skill is the best fit for the k-nearest neighbors. If a fit cannot be found, the skill is placed outside the current area. The centre of the map represents the most connected set of skills, which can be considered as core work skills in this case. These skills include communication, experience, service, business etc. from the result. Furthermore, all the skill blocks next to each other are strongly connected.

The job portals and advertisements are often relatively focused on one type of work. In Kenya, we could see that the software industry and ICT skills are at the core of the entire labor market demanded skills. We could analyze that those are overrepresented in the raw data from the online job portals compared to traditional industries like agriculture, farming and heavy machinery. Also, different portals consist of different types of work; the analyzed portals consist of blue-collar work and middle management job openings, while LinkedIn job openings, for instance, are more related to management, leadership and specialist jobs.

Key insights - TOP 10 Industries, TOP 20 jobs and skills

The tables 1-4 summaries the key findings from the data: top Industries, skills and jobs based on data. From this data analysis, we can show historical trends and current skill needs. We can predict what skills or jobs are down-trending, i.e. what skills and jobs are most likely not to be sought in the near future. Future forecasts, however, are very limited because of constraints in the data and changing vocabulary – AI nor we do not know future job titles, nor the names of future skills. The terminology used to describe skills is constantly changing. No one is searching for ‘large-scale automated data processing,’ which would be appropriate in the 1980s, but the term ‘big data’ popularly describes that concept today.

<table>
<thead>
<tr>
<th>TOP 10 INDUSTRIES</th>
<th>2019 (est)</th>
<th>2018</th>
<th>2017</th>
<th>2016</th>
<th>2015</th>
</tr>
</thead>
</table>
governmental       | 10.604     | 10.624| 5.400| 3.942| 5.843|
marketing          | 5.702      | 7.241| 6.089| 9.320| 8.313|
technology         | 5.668      | 6.096| 5.044| 3.701| 7.078|
healthcare         | 4.647      | 4.281| 3.422| 2.631| 3.373|
media              | 3.200      | 3.436| 2.422| 2.834| 2.820|
customer_services  | 2.945      | 3.330| 4.044| 3.460| 4.590|
food               | 2.774      | 2.431| 2.422| 1.937| 3.244|
hotel              | 2.553      | 1.586| 1.889| 1.841| 2.710|
bank               | 2.179      | 2.502| 2.467| 1.851| 1.751|
education          | 1.600      | 1.770| 1.978| 1.561| 1.889|

Table 1: Top 10 industries during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017)
The arrow showing the trend is a time series estimate. The arrow indicates not only the change from the previous year but also the trend existing over the years. If the trend is not clear, a yellow 'no-trend' symbol is shown. When conducting a trend analysis, we have to be very careful not looking after too short time intervals with too little measures. Predictions can be built only based on non-random changes.

### Table 2: Top 20 job titles during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017)

<table>
<thead>
<tr>
<th>TOP 20 JOBS</th>
<th>2019 (est)</th>
<th>2018</th>
<th>2017</th>
<th>2016</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>officer</td>
<td>↓ 7.252</td>
<td>7.450</td>
<td>8.191</td>
<td>7.852</td>
<td>7.389</td>
</tr>
<tr>
<td>manager</td>
<td>↓ 4.454</td>
<td>4.964</td>
<td>5.360</td>
<td>8.325</td>
<td>5.796</td>
</tr>
<tr>
<td>assistant</td>
<td>↓ 4.034</td>
<td>4.901</td>
<td>7.157</td>
<td>4.621</td>
<td>5.889</td>
</tr>
<tr>
<td>worker</td>
<td>↑ 2.899</td>
<td>2.378</td>
<td>1.753</td>
<td>2.034</td>
<td>1.574</td>
</tr>
<tr>
<td>provider</td>
<td>= 2.336</td>
<td>2.577</td>
<td>2.899</td>
<td>2.537</td>
<td>1.833</td>
</tr>
<tr>
<td>researcher</td>
<td>= 2.294</td>
<td>2.559</td>
<td>2.393</td>
<td>2.182</td>
<td>2.574</td>
</tr>
<tr>
<td>reporter</td>
<td>↑ 2.185</td>
<td>1.649</td>
<td>1.865</td>
<td>1.734</td>
<td>1.685</td>
</tr>
<tr>
<td>accountant</td>
<td>= 2.185</td>
<td>1.649</td>
<td>1.865</td>
<td>1.734</td>
<td>1.685</td>
</tr>
<tr>
<td>clerk</td>
<td>↑ 1.471</td>
<td>1.009</td>
<td>1.326</td>
<td>0.768</td>
<td>1.315</td>
</tr>
<tr>
<td>director</td>
<td>= 1.345</td>
<td>1.261</td>
<td>0.910</td>
<td>1.369</td>
<td>1.213</td>
</tr>
<tr>
<td>programmer</td>
<td>= 1.345</td>
<td>1.495</td>
<td>1.258</td>
<td>1.611</td>
<td>1.556</td>
</tr>
<tr>
<td>marketer</td>
<td>= 1.303</td>
<td>1.360</td>
<td>1.124</td>
<td>1.660</td>
<td>1.361</td>
</tr>
<tr>
<td>developer</td>
<td>↑ 1.092</td>
<td>0.946</td>
<td>0.674</td>
<td>1.044</td>
<td>0.657</td>
</tr>
<tr>
<td>chief</td>
<td>↑ 0.950</td>
<td>0.910</td>
<td>0.719</td>
<td>0.601</td>
<td>0.630</td>
</tr>
<tr>
<td>designer</td>
<td>= 0.840</td>
<td>0.739</td>
<td>0.483</td>
<td>0.507</td>
<td>1.028</td>
</tr>
<tr>
<td>cashier</td>
<td>↑ 0.630</td>
<td>0.378</td>
<td>0.191</td>
<td>0.222</td>
<td>0.389</td>
</tr>
<tr>
<td>offerer</td>
<td>↓ 0.588</td>
<td>0.910</td>
<td>0.899</td>
<td>0.936</td>
<td>0.611</td>
</tr>
<tr>
<td>deliverer</td>
<td>= 0.571</td>
<td>0.685</td>
<td>0.742</td>
<td>0.596</td>
<td>0.417</td>
</tr>
<tr>
<td>trainer</td>
<td>= 0.571</td>
<td>0.568</td>
<td>0.618</td>
<td>0.473</td>
<td>0.667</td>
</tr>
<tr>
<td>secretary</td>
<td>= 0.445</td>
<td>0.441</td>
<td>0.438</td>
<td>0.369</td>
<td>0.574</td>
</tr>
</tbody>
</table>

General job title like “Provider” is a computationally generalized job title found from the data and can be described like cluster/basket words in which “Provider” is connected to several more detailed jobs. This is comparable to job title like “Doctor,” which includes medical doctors, doctor of surgery, doctor of neurology as well as doctor of psychology or doctor of philosophy. All these words differ significantly from each other. This shows the limitation of just using the AI tool since the human efforts are required to validate and analyze the relations for the word. Following is the list of different skills and words related to Provider.
Table 3: Job title relations to the title Provider

<table>
<thead>
<tr>
<th>Provider</th>
<th>2019 (est)</th>
<th>2018</th>
<th>2017</th>
<th>2016</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Solutions_Provider</td>
<td>3.469</td>
<td>3.091</td>
<td>4.078</td>
<td>3.294</td>
<td>5.118</td>
</tr>
<tr>
<td>Service_Provider</td>
<td>3.359</td>
<td>3.783</td>
<td>3.739</td>
<td>4.477</td>
<td>3.671</td>
</tr>
<tr>
<td>Logistic_Provider</td>
<td>2.796</td>
<td>3.296</td>
<td>4.283</td>
<td>2.292</td>
<td>3.519</td>
</tr>
<tr>
<td>Health_Care_Provider</td>
<td>2.327</td>
<td>2.165</td>
<td>1.967</td>
<td>0.966</td>
<td>0.873</td>
</tr>
<tr>
<td>Hts_provider</td>
<td>1.367</td>
<td>1.787</td>
<td>1.522</td>
<td>2.330</td>
<td>1.903</td>
</tr>
<tr>
<td>IT_Service_Provider</td>
<td>1.359</td>
<td>1.504</td>
<td>1.261</td>
<td>0.925</td>
<td>1.312</td>
</tr>
<tr>
<td>Customer_Care_Provider</td>
<td>1.122</td>
<td>1.543</td>
<td>1.444</td>
<td>1.836</td>
<td>1.034</td>
</tr>
<tr>
<td>Software_Solutions_Providers</td>
<td>1.094</td>
<td>1.217</td>
<td>1.178</td>
<td>1.065</td>
<td>1.160</td>
</tr>
<tr>
<td>Medical_Diagnostics_Provider</td>
<td>0.971</td>
<td>0.922</td>
<td>0.733</td>
<td>0.723</td>
<td>0.603</td>
</tr>
<tr>
<td>Travel_Services_Provider</td>
<td>0.747</td>
<td>0.770</td>
<td>0.422</td>
<td>0.296</td>
<td>0.519</td>
</tr>
<tr>
<td>ICT_Provider</td>
<td>0.706</td>
<td>0.822</td>
<td>1.011</td>
<td>0.865</td>
<td>1.051</td>
</tr>
<tr>
<td>Day_Shift_Provider</td>
<td>0.612</td>
<td>0.596</td>
<td>0.333</td>
<td>0.540</td>
<td>0.367</td>
</tr>
<tr>
<td>Security_Provider</td>
<td>0.531</td>
<td>0.457</td>
<td>0.333</td>
<td>0.511</td>
<td>0.300</td>
</tr>
<tr>
<td>Data</td>
<td>0.482</td>
<td>0.435</td>
<td>0.450</td>
<td>0.239</td>
<td>0.422</td>
</tr>
<tr>
<td>Data</td>
<td>0.496</td>
<td>0.404</td>
<td>0.261</td>
<td>0.275</td>
<td>0.329</td>
</tr>
<tr>
<td>Technology</td>
<td>0.441</td>
<td>0.461</td>
<td>0.294</td>
<td>0.487</td>
<td>0.430</td>
</tr>
<tr>
<td>Ethics</td>
<td>0.408</td>
<td>0.348</td>
<td>0.244</td>
<td>0.227</td>
<td>0.439</td>
</tr>
<tr>
<td>Design</td>
<td>0.408</td>
<td>0.213</td>
<td>0.267</td>
<td>0.306</td>
<td>0.105</td>
</tr>
<tr>
<td>Maintenance</td>
<td>0.400</td>
<td>0.478</td>
<td>0.494</td>
<td>0.390</td>
<td>0.451</td>
</tr>
<tr>
<td>Internet</td>
<td>0.347</td>
<td>0.391</td>
<td>0.639</td>
<td>0.692</td>
<td>1.143</td>
</tr>
</tbody>
</table>

Table 4: Top 20 skills during 2015-2019. Values are group’s shares (%) of yearly demand in general. (1% is 215 jobs in 2019, 230 jobs in 2018, 110 jobs in 2017)

Table 4 analyses the Top 20 skills from the data. Top 20 skills list includes both soft and hard skills as AI does not differ these skills. This pilot study focused on providing Top 20 lists. It would
be also interesting to explore the data from different perspectives; e.g. which skills are disappearing, or which are in demand constantly. This data analysis is doable but requires further data science work.

Kenyan software and ICT industry skills demand

When focusing the labor market analysis on one thematic area (like software industry and ICT) over a shorter time frame (January 2019-June 2019 in this case), it is easier to understand the skills sought in those areas. Because of the data complexity, the more data we include into maps, the more non-optimal relations occur in the map. Non-optimal relations are not incorrect, but they make the visualization more difficult to read. The Figure 6 skills map presents the skills demand in Kenya in the following domains: SOFTWARE, ICT, PROGRAMMING, JAVA, PYTHON, SCRIPT, JAVASCRIPT, SQL, DATA, NETWORK, SERVER, BACKEND, FRONTEND, CLOUD_COMPUTING, CLOUD, HTML, and AGILE.

At this this domain, experience, knowledge, technical, technology, data as well as business, communication, customer service and project skills seem to be most wanted.
Table 5: TOP 20 skills in Kenyan software & ICT industry skills demand

Even though this data focus on software and Information Technology (IT) industry, it looks like business and soft skills seem to be the top skills rather than technical skills. But if you look at the graph on technical vs. software skills, it looks like still more technical skills are in demand.

Curriculum skills maps and gap analysis

This project was also an experiment to take advantage of new methods to understand the phenomenon - how to machine-read curriculums. The challenge of this method was the variety of curriculum descriptions. Universities each use their own vocabulary, concepts in different ways and degrees - there is no uniform way of writing the curriculum. Also, each university has its own way to share the curriculum on their website. For example, University of Nairobi provided the course descriptions on their website as an html structure and MOI University had the information both on their website as well as pdf documents. These different formats make it challenging for a machine to read as you cannot use the same method. At the same time, students may also have a hard time understanding each curriculum and making informed decisions about the selection of the schools and courses.

Headai’s AI machine-read the B.Sc and M.Sc curriculums in computer science program in two universities from Kenya. AI builds a skills map/Digital_Twin on curriculum as a model on the chosen educational offering.

In the curriculum map, the basic unit is one skill, which is represented by a block. The more the skill words are used, the deeper the green of the block becomes. The figure is clustered with an applied k-nearest clustering algorithm, i.e. the origin in the middle of the figure is the skill that has the most connections to other skills. The next connected skill is the best fit for the k-nearest neighbors. If a fit cannot be found, the skill is placed outside the current area. The center of the map represents the most connected set of skills, which can be considered as core skills being taught in this case.
For instance, among skills taught at the university of Nairobi’s computer science program, project, knowledge, software, problem solving, communication, design, research, technology, ICT, testing are the top 10 skills taught.

On the other hand, in Moi University, as the Figure 9 shows, science, communication, manager, publishing, operating, problem solving, processing, service, technology, electronics are the top 10 skills words used in the computer science curriculum.
These skills maps give teachers, education designers and administrators an immediate overview of how they use skills words in the curriculum and education. If teachers teach all the skills demanded by the labor market but document them in the curriculum with different words, it will not help students identify what kind of skills they are learning. Students also cannot specifically present the skills that they gain using the vocabulary used by the employers. The skills map can also provide an overview on how much of the universities’ curriculum focuses on technical skills compared to business skills. The Figure 9 and 10 show the comparison between Moi University and University of Nairobi on their computer science program. Moi University has more business and soft skills compared to University of Nairobi.
Similarities and differences between skills demand and educational offering

With the knowledge from the labor market and curriculum offering, it was possible to create skills maps that compare the demand and supply of skills. The visualization shows the possible similarities, differences and overlaps between educational offering and skills demand. In the Figure 11, the yellow blocks represent overall labor market demanded skills that are trained at the university. The red blocks represent skills demanded that are not taught at the university. The green blocks represent the skills taught at the university but are not demanded in the labor market. The original plan was to conduct a virtual workshop with two universities. However, only Moi University responded to the request.

Moi University

Figure 11: Skills taught at Moi university in software and ICT. Similarities and differences between skills demand and educational offering.
The visualized skills maps enable different users, including teachers, managers, lecturers, career coaches and students, to perform searches related to various skills demand. The different stakeholders can utilize visualized skills maps in planning and collaboration. These maps aim to produce value in creating new insights by anticipating social changes, yet-unrecognized skill needs and skills clusters.

Headai identified four different skill gap categories for MOI University.

In category 1, under professional software developer jobs, the following are identified as example skill gaps: data science, big data, python, machine learning, SQL, server, git, cloud computing, amazon web services, engineering, technical, and optimization. These skills gaps seem to be related to the latest digital skills such as big data analytics and machine learning as well as the use of cloud computing. Moi university plans to incorporate current topics of Computer Science including: Internet of Things, Machine learning, Big Data, edge computing and cloud computing.
In category 2, under industry platforms and tools, the following skills are identified as example skill gaps: telecom, mobility, ISO 9001, platforms, certifications, industry, SAP, IBM, Android, Java, and Oracle. These gaps seem to exist in industry standards as well as enterprise package software. While the university seems to be not offering this type of course, through university-industry partnerships, special short courses can be provided by the industry which offers those IT professional certificates. Moi university plans to incorporate the current computing platforms especially Amazon web services and object-oriented programming using Java to enable students develop android applications.
In category 3, under financing and business-related ICT skills, the following skills are identified as example skill gaps: financing, analytical, evaluation, customer, data-science, planning, reporting, accounting, auditing, stocks, contracts, logistics. Since the AI tool presents one skill word per block instead of a combination of words, it is not completely intuitive in terms of the skills gap, but it is possible to guess. In finance and business-related ICT skills, there seems to be a gap in teaching business relevant ICT skills such as data analytics as well as specific financial knowledge and business transaction skills. Moi university intends to have entrepreneurship, marketing and business intelligence courses in the curriculum or even have short courses. Moi university considers data science to be also an important component of computer science.
In category 4, under basic ICT skills, the following skills are identified as example skill gaps; artistic, editing, Microsoft Office, Excel, Photoshop, and Illustrator. These skills gap seems to exist in a secretory and administration type of jobs as well as creative jobs. Moi university intends to incorporate photoshop and illustrator in its COM 110 and COM 111 of the current curriculum. In computer graphics course, they will provide students with design software including: Adobe and illustrator.

**Feedback from MOI University**

The project aimed to share the analyzed data and results with universities as well as get their feedback on the method. At the virtual workshop with MOI University, the World Bank and Headai discussed the potential use of data analysis for university curriculum development and optimizing the future supply of education to match employers’ skill needs. University representatives were very impressed with the approach and process described and saw how it helps the university to project the future skills demand, review the curriculum, and engage with the labor market. The approach and method would inform the review process of the curriculum, keeping the curriculum up-to-date and aligned with the labor market demands. It is also useful for the university to validate the impact of the curriculum, as well as ensure their graduate and post-graduate students succeed in the labor market. There was an interest in scaling the model into other departments and faculties, schools of arts and school of engineering. There was also a demand for assessing the skills level of students and teachers, which the World Bank is planning to cover during the next phase of the pilot.

Moi university considers the following points as their key action points:
1. **Use of online courses to supplement curriculum gap**: Computer Science is a very dynamic field and no curriculum will ever match this dynamism. There is a need to encourage students to subscribe to online courses and the department can subscribe for a few courses for them.

2. **Learners should be equipped with current computing fields** including the Internet of Things, Edge Computing, Cloud Computing, and Big Data. This can be done through short courses, hackathons, and even workshops and symposia.

3. **Partner with High Profile computing companies** including IBM, Amazon, Microsoft, Google to offer students internships and attachments to supplement the curriculum.

4. **Encourage multi-disciplinary approaches** to systematically develop students’ skills: when students are doing their projects, for instance, have computer science, applied statistics and actuarial science students working together to solve a problem that cuts across their areas.

5. **Empower Teaching staff with current computing skills** This involves establishing a link between the teaching staff and the industries on the current computing trends.

---

**University of Nairobi**

![Image: Skills taught at University of Nairobi in software and ICT. Similarities & differences between skills demand and educational offering]

Headai analyzed and identified four different skills gap categories for University of Nairobi.
In category 1, under professional software developer, the following skills are identified as example skill gaps: git, Python, business intelligence, data science, big data, operations, tools, edge. In both universities, git, python, data science, and big data came up as common gaps, but Moi university had extra gaps in infrastructure relevant skills such as cloud computing, SQL, and server. In fact, University of Nairobi has taught skills related to those identified as gaps. However, they describe the skills differently in the curriculum compared to the labor market such as artificial intelligence, statistics, data mining, cognitive, and problem solving (green words). If the university uses similar words used in the online job advertisements in their curriculum and course description, the gap will look smaller.
In category 2, under industry platforms and tools, the following skills are identified as example skill gaps; certifications, ISO-9001, Amazon web services, IBM, Android, Oracle, mobility, business intelligence. These are almost the same as the ones from Moi University. However, Moi university had a little more skills gap especially in the coding language like Java.
In category 3, under financing and business-related ICT skills, the following skills are identified as example skill gaps: tools, big data, operating, business intelligence, data science, marketing, customer, bank, accounting, financing, negotiations, payments, contracts. Compared to the gap identified in Moi university’s curriculum, University of Nairobi has similar gaps in data science and big data analytics. In addition, they also have the similar skills gaps in financial technical knowledge such as accounting, banking and financing. University of Nairobi seems to have more gaps in other areas of business skills, including marketing and operation.
In category 4, under Basic ICT skills, the following skills are identified as example skill gaps: excel, editing, artistic, Photoshop, Illustrator. These are also similar to the ones from Moi University. Both universities seem to lack design-relevant basic ICT skills.

Summary and recommendations for the future

The aim of this project was to build the understanding on Kenyan skills demand based on online job portal data and Kenyan ICT educational offering based on the curriculums from two universities. All experiments were based on publicly available data, i.e. job advertisements and online course descriptions. The job advertisements were dominated by ICT jobs, which is not a problem when focusing on skills demand and offering in ICT and software industry. However, this indicates that such studies would be difficult to include other sectors such as agriculture, forestry, and informal sector.

The main findings of the project were that Kenyan universities are currently providing education that mostly matches the labor market demand. Universities can use the outcomes of this project to finetune the educational offerings to better serve specific high-demand areas. At the same time, it is important to recognize that universities should not change the core components of the curriculum. Modifying the language used in curriculum and courses could also enable students to clarify what skills they are learning by taking specific courses which they can later describe on their resume appropriately.
From the methodological and technological point of view as well from the MOI University feedback, this study validates that the method and approach can bring significant added value and positive impact for the universities. It enables a faster update cycle for curriculum development and an accurate and up-to-date view of the skills that are sought in the labor market. While this study showed the potential of the AI tool in labor market analysis and curriculum gap analysis, human efforts were still needed to analyze the data further to make an appropriate implication and show the results in an easy-to-understand manner. Thus, a combination of human and AI is needed to understand the skills gap. Also, it is crucial to understand that the current tool works as an analysis software for research scientists to help them process critical data as SPSS or other statistical software do. In order to serve wider range of end-users and practitioners, presentation of the data analysis needs to be more non-technical and user-friendly using infographics and data visualizations for example.

From the research and development point of view, it is important to continuously monitor and refine the educational offering and training based on online job posting data from the portal. Without continuous AI training process, AI would also not be able to recognize the new skills and topics. Furthermore, the language used in job advertisement evolves: domain once called electronic data processing turned to IT and ICT. Even though electronic data processing is a valid word, no one will use it today, especially in job advertisements.

Suggestions for further opportunities

- **Provide students with information on how to select courses wisely.** Based on the skills demand and course offering information, a system can be built to offer individual learning and career path options for each student. This would help students achieve their career goals. The system can also show how the selection of courses at the university may affect their potential job opportunities.

- **Use of online courses and short courses to supplement the curriculum gap.** Since the curriculum update cycle does not necessarily meet with the skills demanded in the labor market, one way to solve this issue is to offer short courses or use online courses from MOOCs. This is also aligned with the action points that Moi university suggests.

- **Analyze the skills and competencies of the faculty members and identify their skill needs for in-service teacher education.** Though the course offering at two sample universities covers the skills demand well, continuous changes in industries and new technologies make an upskilling of faculty members a critical component for universities. Moi university suggests that the linkage between faculties and industry can be also made to improve the current computing skills of the faculties.

- **Develop a toolkit of different data infographics and visualization templates for labour market and curriculum data analysis.** This would help scale the method and share the results for a wider scale of end users, practitioners and decision makers in a user-friendly and easy to consume format.

- ** Develop training materials for professionals to analyse, interpret and apply the results in various contexts.** The methodology used for this analysis is similar to statistics. Thus, without training in applying k-means clustering, people would not get much out of it, but statistically trained people find the analysis to be useful. Thus, training materials are needed to educate more people to get the most out of this analysis.

- **Provide opportunities to learn/benchmark other universities’ curriculum.** Continuous improvement of curriculums and course descriptions require benchmarking among
universities. It would be useful for the university management to learn about the examples from other universities to improve their curriculum. More detailed future forecasts can be possible if other data sources could be used. For instance, public investment announcements and governmental foresight reports could be used as possible data sets to help improve the quality of future forecasts.

- **Enable cross-disciplinary approaches to systematically develop skills of faculty and students.** Since more combined skillsets from different disciplines would be needed in the future, it would be useful for students and faculties to learn in an interdisciplinary approach while deepening one or two specific subjects. Moi university aligns with this approach.

- **Scale this method to several domains in several countries.** We should understand in more detail country-wise differences in skills demand and educational offerings. This requires not only taking this method to new countries, but also taking new domains into a study. In Kenya, ICT and software industries were overrepresented in online job portals, but we cannot be sure if it is the same in other African countries. By providing wide understanding on skills demand, we can help universities to make right decisions when developing the future curriculums and courses. A natural continuation for this project could be collaborating with data from Finnish universities that have used Headai methodology and technology. Headai data analysis together with Universities of Applied sciences in Finland has found similar ICT related skills demand as the ones found in this project. In addition to technical skills, Finnish labour market data seem to emphasize more soft skills like teamwork, communication, quality, courage, willingness to learn etc. Also, a special interest of university has been on new combination of skills: e.g. healthcare and data-science, data and customer work, and network and control.

- **Bring up the discussion around open data to online job portal providers and universities.** Under this study, we could analyze the public data from both job market and universities. Without having programming interfaces like API to online job portals data, we needed to crawl, scrape or mine the data from online job portals with non-standard means. This increases the project costs and durations. If we could ask the online job portal companies to provide us with an access to their job posting data, data collection would become easier, cheaper and faster. An open data model would also open new business models for job portals as well.
References


Sitra. (2017) Artificial intelligence shows what Finland can do and a positive CV reveals the hidden talents of young people