Building a Needs-based Curriculum in Data Science and Artificial Intelligence: Case Studies in Indonesia, Sri Lanka, and Thailand

Chutiporn Anutariya[1], Marcello Bonsangue[2], Taufik F. Abidin[3], Matthew Dailey[1], Katerina Fraidaki[4], Tiago Gomes[6], Felienne Hermans[2], Suratsavadee Korkua[5], João L. Monteiro[6], Nugraha P. Utama[7], Sofia Pereira[6], Amalka Pinidiyaarachchi[8], Sandro Pinto[6], Jan N. van Rijn[2], Opim Salim Sitompul[9], Wararat Songpan[10], Chitsutha Soomlek[10], Frank Takes[2], Suzan Verberne[2], and Chitraka Wickramarachchi[11]

[1] Asian Institute of Technology, Thailand
[2] Leiden University, the Netherlands
[3] Universitas Syiah Kuala, Indonesia
[4] Athens University of Economics and Business, Greece
[5] Walailak University, Thailand
[6] University of Minho, Portugal
[7] Institut Teknologi Bandung, Indonesia
[8] University of Peradeniya, Sri Lanka
[9] Universitas Sumatera Utara, Indonesia
[10] Khon Kaen University, Thailand
[11] University of Sri Jayewardenepura, Sri Lanka

Email corresponding authors: chutiporn@ait.ac.th, m.m.bonsangue@liacs.leidenuniv.nl

Abstract

Indonesia and Thailand are middle-income countries within the South-East Asia region. They have well-established and growing higher education systems, increasingly focused on quality improvement. However, they fall behind regional leaders in educating people who design, develop, deploy and train data science and artificial intelligence (DS&AI) based technology, as evident from the technological market, regionally dominated by Singapore and Malaysia, while the region as a whole is far behind China. A similar situation holds also for Sri Lanka, in the South Asia region technologically dominated by India. In this paper, we describe the design of a master's level curriculum in data science and artificial intelligence using European experience on building such curricula. The design of such a curriculum is a nontrivial exercise because there is a constant trade-off between having a sufficiently broad academic curriculum and adequately meeting regional needs, including those of industrial stakeholders. In fact, findings from a gap analysis and assessment of needs from three case studies in Indonesia, Sri Lanka, and Thailand comprise the most significant component of our curriculum development process.

Keywords: Curriculum Design; Higher Education; Data Science; Artificial Intelligence.

1 Introduction

The use of data science and artificial intelligence (DS&AI) techniques is rapidly growing in several sectors around the world, e.g., in the health care (Jiang, Jiang, Zhi, et al., 2017), transport (Abduljabbar, Dia, et al., 2019) and financial sectors (Gomber, Koch, and Siering, 2017). This revolution is accelerated by advances in data collection, storage and analysis, efficient algorithms, and an increasing computer processing power. Although the United States, China, and Europe (Carriço 2018) are the frontrunners in developing such DS&AI methods, these world-wide scientific advances also have a major impact on other countries, such as Thailand, Indonesia, and Sri Lanka. As the fields of DS&AI have the potential to contribute positively to the economic and social climate of those countries, there seems to be a great need for training professionals in this area (McKinsey Global Institute, 2017b).

Thailand, Indonesia, and Sri Lanka are middle-income countries within the South-East and South Asia region. While DS&AI based technologies are gaining traction and becoming strategic priorities, these countries are still falling behind regional leaders in the technological market, such as Singapore and Malaysia, while the region as a whole is far behind China (McKinsey Global Institute, 2017a). Thailand, Indonesia, and Sri Lanka

have well-established and growing higher education systems, increasingly focused on quality improvement. However, they fail in educating people who design, develop, deploy, and train DS&AI based technology, as evident from the acute shortage of data analysts in public and private companies (McKinsey Global Institute, 2017c, Mathur, Aneja, et al., 2019).

These trends make clear that development of talent and capabilities is needed if DS&AI are to reach their full potential throughout the region. The success of political and economic initiatives by governments in these countries to transform the current export-based economy into an innovation- and knowledge-based economy will critically depend on the readiness of the labour force for ICT in general and data-driven technologies in particular. Economic and social development increasingly depend on innovation (Bundy, 2017). Universities have an important role in driving innovation and development. They can do so both through their role in carrying out research and development and by training workers for the knowledge economy. Higher education is not only playing an important role in human resource development, but also in bringing students to companies through internships and in the regional workforce education via professional training.

In this paper, we investigate the great demand of employees in the field of DS&AI in Thailand, Indonesia, and Sri Lanka. In particular, we look at the requirements that small-, medium-, and large-sized companies, both public and private, have on their next generation of workers. These requirements impose new challenges for higher education in computer science across these countries: maintaining and improving education quality (Wittayasin, 2017); and improving the relevance of curriculum and instruction at a time of rapid change in labour market needs. Extraordinary effort is demanded to diversify curricula and to develop, for example, a needs-based curriculum for a master's degree in the fields of DS&AI. These needs are vital for companies to grow larger and keep pace with the advancement in the rest of the world, more importantly in USA, China, and Europe. The design of a novel curriculum in DS&AI is a nontrivial exercise because we have to take into account a constant trade-off between having a sufficiently broad academic curriculum and adequately meeting regional needs, including those of industrial stakeholders.

This paper reports the work compiled using the input from three different activities in the context of the *Erasmus+* project on *Curriculum Development in Data Science and Artificial Intelligence*, registered under the reference number *599600-EPP-1-2018-1-TH-EPPKA2-CBHE-JP*. We proceed as follows. In Section 2 we identify the existing academic curricula, in Section 3 the labour needs from industry, and in Section 4 the gaps that need to be fulfilled by a new curriculum in DS&AI. The curriculum development is described in Section 5. Finally, we draw some conclusions in Section 6.

2 Identifying existing academic curricula

Through desktop research and using a common template, we have collected data from 35 master's degree programmes in DS&AI in Thailand, Sri Lanka, and Indonesia. Many of the programmes we collected data from are generic master's in computer science and contain only few courses related to data science or artificial intelligence. Only 13 out of 35 programmes are directly related to data science or artificial intelligence. Of these, only two are full-fledged programmes on artificial intelligence. The main reason is because in Asia computer science education programmes are often developed in a generic way, e.g., based on ACM curricula recommendations. Most common programmes are therefore named Computer Science, Information Technology, Information System, and Computer Engineering. These programmes often include several elective tracks, such as hardware and computer architecture, communication and networking, computer systems or software. However, there are almost no programmes specific to DS&AI. As a result, students are only allowed to follow a limited number of elective courses within their track, limiting their dedicated knowledge on data science or artificial intelligence.

In Thailand, the fields of DS&AI are currently part of the national strategic plan, called Thailand 4.0, which is a 20-year strategy to accelerate the country's development from upper-middle-income country to high-income country status through digital transformation of manufacturing. Therefore, the development of study programmes in DS&AI are encouraged by the government, although only four dedicated master's degree programmes in Data Science (and none in Artificial Intelligence) are currently available. A related major

challenge comes from the northern and north-eastern regions of Thailand, because of the limited number of experts who are eligible to teach and supervise research in master's and doctoral programmes.

In Indonesia, the name of the study programmes in the computing field must be as specified in the ACM Computing Curricula: Computer Science, Information Technology, Information Systems, or Computer Engineering. Due to the above naming regulations, we were able to identify only three specific master's programmes in DS&AI, while several courses are typically available under existing computer science, information technology, and Information systems curricula.

The situation in Sri Lanka is very similar to the one in Thailand, with a government pushing towards a new digital revolution in industry, but with a not well developed higher education programmes specialised in DS&AI to absorb the increasing request from the labour market in these areas.

To summarize the curricula of the existing programmes in areas related to DS&AI, we divided the course types into three groups: core, elective, and supportive course. The core courses are the mandatory courses in the curricula. The elective courses are additional ones that can be selected by students among several alternatives. The supportive courses are related to the basic knowledge which may be the pre-requirement for admission or a preparatory course for the programme. Skills such as communication and presentation, and DS&AI ethics are not always present as full-fledged courses, but incorporated into the programme of other core, elective or supportive courses. The most popular topics among the core courses in all programmes we analysed are data management, data mining, data warehousing, machine learning and research methodology. There are various elective topics which are hard to categorize because of a non-uniform naming. There is only one topic which is taught in all programmes: data and text analytics. Finally, all programmes make available three groups of supporting knowledge: programming, algorithmics, and mathematics/statistics.

3 Identifying labour needs from industry

Across South-East and South Asia regions, talent shortages in DS&AI—caused by gaps and mismatches between labour market needs and the available supply of educated and trained professionals—constrain economic growth (Skinner, Saunders and Beresford, 2004). These shortages prevent companies from scaling up, meeting demand in new locations, and launching new products and services. The talent shortage will have a direct impact on their competitiveness and productivity. We explored the current needs that companies in Thailand, Indonesia, and Sri Lanka have for DS&AI experts. We gathered information systematically through a questionnaire sent to several companies in the above countries, asking to quantify their needs in DS&AI job positions and internship, and which knowledge/skills they require or would ask for professional trainings. We differentiated between the needs of the companies by their size in terms of employees (small: < 50, medium: 50-249, and large: ≥ 250 employees), and we covered several sectors including both public and private companies. We received a response from 59 companies: 23 from Thailand (no medium company responded), 26 from Indonesia, and 10 from Sri Lanka. Here, we consider labour needs not only concerning knowledge, but also those that can occur in skills (programming) or practice (tools).

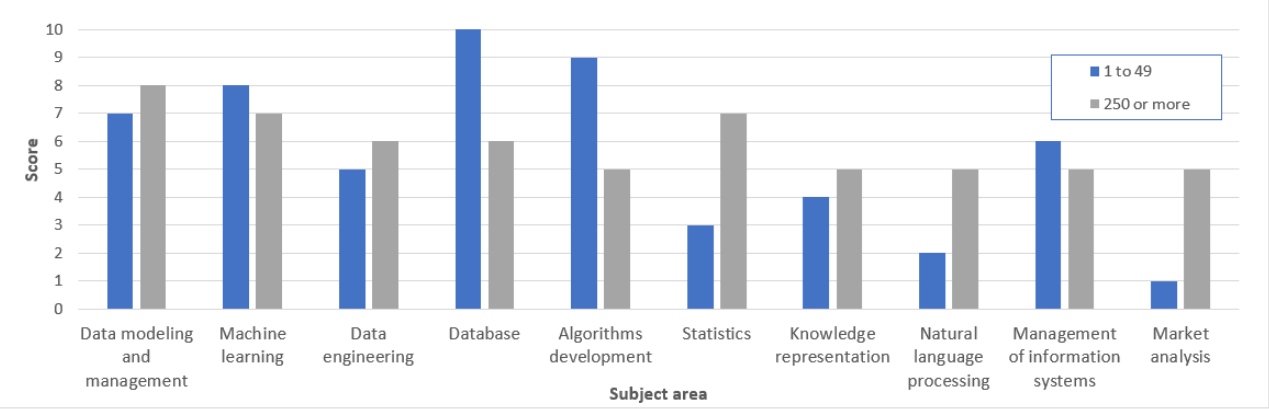

Figure 1. Subject areas with high demand in Thailand for internship placement of companies with different sizes.

In Thailand, an analysis reveals the highest demand for jobs as data scientist, AI/data analyst, software engineer, and business analyst, respectively. The demands of job positions follow a similar trend for small- as well as large-sized companies. Companies require interns to have a strong background in programming (Python, R, SQL, Anaconda, Tensorflow, Genetic Algorithm, and NoSQL), to be fast learners, to have basic skill to complete the end-to-end workflow (public cloud computing platform, SAP, Microsoft Office 365, Power BI, Microservice (Docker for deployment project)), and to have knowledge on several subject areas as illustrated by Figure . The priority on the knowledge required lies on data modelling and management, databases, machine learning, data engineering, and algorithms development, and statistics. The most important subject areas for professional trainings are data modelling and management, machine learning, data engineering, and knowledge representation.

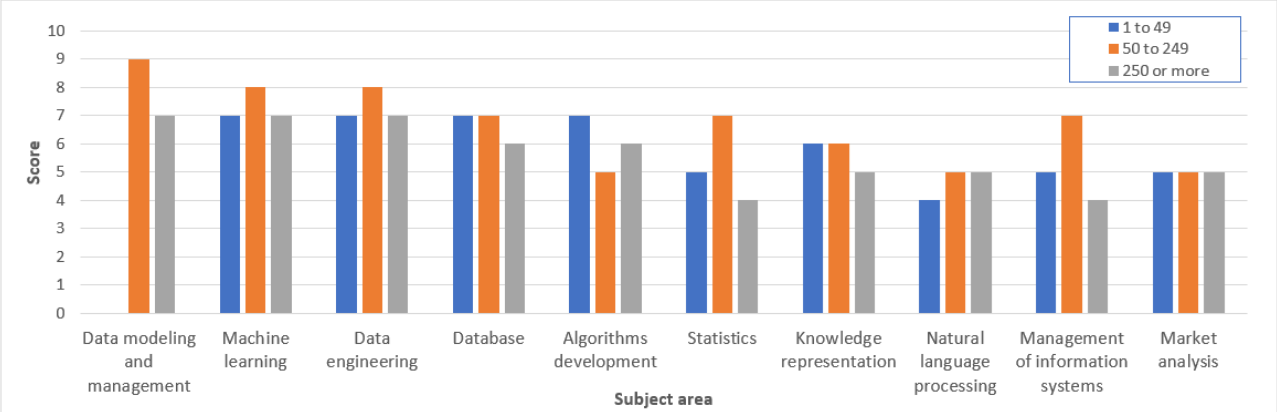

Figure 2. Subject areas with high demand in Indonesia for internship placement of companies with different sizes.

In Indonesia, the highest demand for DS&AI job positions includes data scientist, software engineer, and artificial intelligence/data analyst, respectively. The latter are especially demanded in large-sized companies. All companies we surveyed require interns to have knowledge on data modelling, machine learning and data engineering. In addition, small-sized companies require knowledge on databases and algorithms. Natural language processing is felt in general as not necessary. This is surprising since text analytics is the only course that is taught in all existing programmes (see section 2). Maybe companies do not realize that text analytics is an important subfield of natural language processing. Interestingly, also statistics is one of the knowledge areas that is not really required from interns by large companies. Figure 2 summarizes the average score given by the companies on the required knowledge. Concerning programming languages, required knowledge covers Python, Tensorflow, R, SQL, Java, or any web-based languages, while that for statistical and data visualization tools includes Hadoop, D3.js, Microsoft Power BI, SAP and similar. The needs for training courses cover topics on data modelling and management, machine learning and data engineering. However, the needs for professional courses on statistics, natural language processing and management of information systems are considered, on the average, less interesting.

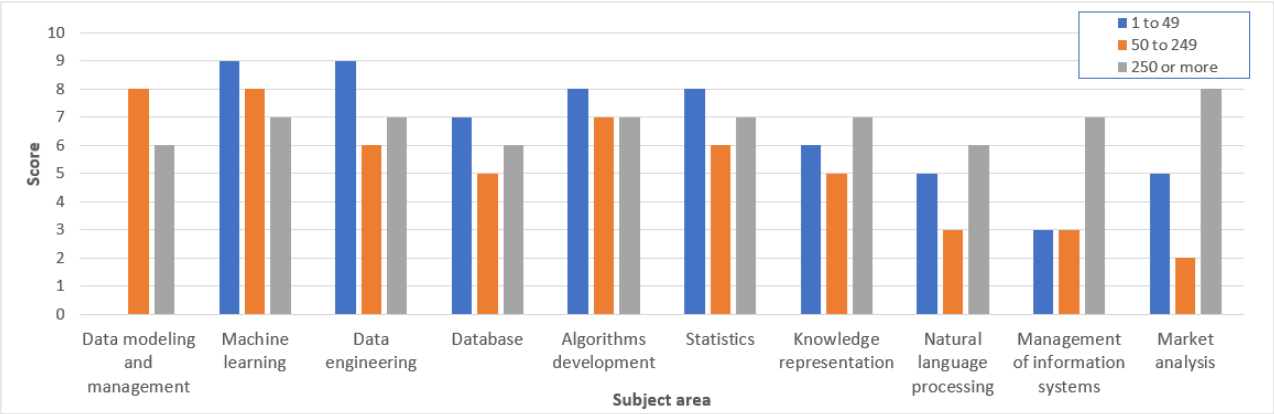

Figure 3. Subject areas with high demand in Sri Lanka for internship placement of companies with different sizes.

In Sri Lanka, we received a response from only 10 companies, 4 small-sized, 4 mid-sized, and 2 large ones. They were all private companies, and all positioned in an ICT related area. Overall, the highest demands for DS&AI job positions are machine learning engineer, data scientist, and artificial intelligence/data analyst, respectively. Students are required to have knowledge of programming language such as Python, TensorFlow, R, SQL, and to be able to use Matlab, Hadoop, and Git. There are large differences between what is considered important as knowledge of the interns, depending on the size of the company. Large companies prefer interns with knowledge of market analysis, whereas small- and medium-sized ones prefer knowledge of machine learning and algorithms, and consider market analysis as the least important subject. Other topics that are considered important are data modelling and management and data engineering. Figure 3 summarizes the findings. The most important subject areas for professional trainings are data modelling and management, machine learning, and data engineering. Large companies find algorithms also very important. The least important subject areas are databases, statistics and market analysis, probably because the employees already have knowledge on these subjects and do not need extra training.

4 Identifying the gaps

In this section, we explicitly identify the gaps to be addressed in the development of master's degree programmes in DS&AI in Thailand, Indonesia and Sri Lanka. With this goal, we organized in each country a focus group composed of at least 4 academics, 3 students and 3 representatives from ICT companies. This composition guaranteed that all relevant stakeholders provided their input in understanding what is needed by a new education programme in DS&AI. We investigated the following issues:

1. Identification of subject areas related to DS&AI that are most in need in the country (and wider region).
2. Identification of a set of skills that the curriculum should develop and promote.
3. Teaching and learning processes that are appropriate for the curriculum (project-based learning, professional certifications, trainings, practical and industrial projects, workshops, internships, research, theses, etc.).
4. Required resources, facilities, tools, as well as support that the universities should provide (what we lack or need).
5. Any concerns and opinions regarding internships and job opportunities that the curriculum should provide.

Next, we briefly summarize the results of the discussion fora in Thailand, Indonesia, and Sri Lanka. The teaching and learning of the DS&AI curriculum must be able to meet both the needs of the industry and the interest of students. Teaching and learning should lay down the foundation of core knowledge that is sufficient for learning DS&AI, including (1) mathematics, logic, and statistics, (2) programming and computational thinking, and (3) data management. The need for knowledge of business analytics and machine learning depends on the student group characteristics and the objectives of the students. Therefore, it is preferred to have different tracks in the curriculum ranging from (1) applying data analytics techniques to solve problems, analysing data for the industry, and (2) developing deep knowledge in DS&AI, which can enhance the higher performance for the particular industry.

The laboratory equipment should be prepared and ready to use for the most effective learning. The tools used in teaching and learning should be in line with those commonly used in the industry, including Google cloud, Hadoop, Python, and Tensorflow. Students should have the skills to use those tools that can lead to the product development, not just a tool that running or testing as experiments. The courses must be built in collaboration with the industry in applying the real case study for teaching or for students to experiment. Students should practice an internship in a company, or work on industrial projects.

Finally, students should not only gain hard skills through courses and internship training, but also soft skills. The soft skills that are suggested to be developed include: (1) communication skills, (2) analytical skills/ empirical skill, and (3) presentation skills, both in terms of how to present and how to use visual tools to effectively communicate the results of data analysis to the users.

5 Developing the curriculum

In agreement with the international views on higher education, the objective of a master's degree programme in DS&AI is to provide students with a suitable basis for a further career, both in research as well as in industry and society in general. As such, it provides the student with the specific knowledge and abilities, exemplified in the form of a master's degree that allows graduates access to a PhD programme in DS&AI and related disciplines. Also, it prepares graduates for a position in which they can earn their own subsistence, for example by working in knowledge-intensive companies. Our starting point in the development of the curriculum was to take into account the needs and gaps described in the previous sections as well as to provide graduates with a basic understanding of all the key areas of DS&AI, and an advanced understanding in some of these key areas.

The programme has a duration of two years, and includes core courses, electives, an internship, and a thesis work. The core courses define the key areas of the curriculum in DS&AI, and are therefore present as a dedicated mandatory course in the programme. From the data we collected, we arrived at the conclusion that there is a general, and largely worldwide (Quality Assurance Netherlands Universities, 2015), consensus that our proposed five core courses below define the key areas of DS&AI:

- Artificial intelligence: This course introduces the students to fundamentals of Artificial Intelligence, in particular to several techniques on planning and decision procedures ranging from precise to uncertain and temporal reasoning with applications to intelligent agents.
- Business Intelligence and Analytics: Business intelligence is a process of analysing business data to obtain business insights and actionable intelligence and knowledge, to support better business decision making and capture new business opportunities. This course gives the students an understanding of the principles and practices of business intelligence and data analytics to support organizations in conducting their business in a competitive environment.
- Computer Programming for DS&AI: This course is a laboratory course that provides students with the computer programming background necessary for preparing, manipulating, analysing, processing, and visualizing data sets as well as for building-data-driven predictive models
- Data Modelling and Management: The course emphasizes on emerging data models and technologies suitable for managing different types and characteristics of data. In particular, the focus is on developing skills for analysing, evaluating, modelling, and developing database applications with concerns on both technical and business requirements.
- Machine Learning: This course introduces the fundamentals and applications of machine learning. Students learn to design, implement, and evaluate intelligent systems incorporating models learned from data.

The above set of core courses is not intended to constitute the complete master's programme. Programme flexibility and adaptability to the regional needs is obtained through additional elective courses. Within the *Erasmus+* project on *Curriculum Development in Data Science and Artificial Intelligence*, we have developed the following 11 elective courses:

- Natural Language Processing
- Computer Vision
- Distributed Systems
- Human Computer Interaction and Information Visualization
- Knowledge Representation
- Multicriteria Optimization and Decision Analysis
- Nature Inspired Computing
- Recent Trends in Machine Learning
- Social Network Analysis
- Software Development and Project Management
- Spatial-Temporal Data Analysis

Core and elective courses cover all aspects of data science intended as the extraction of useful domain knowledge from large, complex data sets. Core courses like *Machine Learning* or *Data Modelling and Management* provides the critical knowledge needed for decision makers based on statistically significant patterns in data. Business applications are treated in courses like *Business Intelligence and Analytics*, while the detection of anomalies in patterns of behaviour is developed in the electives *Social Network Analysis* and *Spatial-Temporal Data Analysis*. Finally, the course *Distributed Systems* treats, for example, the analysis of services in cloud computing.

The curriculum we developed covers the broad field of artificial intelligence as the theory and development of computational systems able to perform tasks inspired by human intelligence. We basically develop on three main views of the field of Artificial Intelligence:

1. *Symbolic AI* – consisting of knowledge representation, logic reasoning, inductive logic programming, precise planning, and decision procedures;
2. *Statistical AI* – consisting of decision networks, probabilistic programming (including Bayesian program synthesis), computer vision (activities and image recognition as well as machine vision), natural language processing, and machine learning; and
3. *Subsymbolic AI* – consisting of distributed artificial intelligence (including agent-based modelling, swarm intelligence and multi-agent systems), distributed computing, natural computing, multi-objectives optimization, and autonomous systems.

It is not a surprise that machine learning plays a pivotal role in the curriculum, as, today, it can be considered the most widely applied discipline within Artificial Intelligence. Machine learning comes in the curriculum in at least three "flavours": supervised learning, unsupervised learning, and reinforcement learning. Supervised learning refers to methods for learning a model from a set of training data containing both input and output results. Regression and classification are typical examples of supervised learning techniques. Unsupervised learning consists of building a model from a data set, via for example visualization or clustering. Reinforcement learning is about taking actions in an environment to maximize some notion of cumulative reward, as for example used in robot navigation and games (including chess, Go, and business games).

As part of the programme, students are required to get few months placement with a company before graduating. As students participate in internships, they gain real work experience, which not only benefits employers, but also provides a valuable complement to university education and an input to job counselling (Vairis, Loulakakis, and Petousis, 2014). All knowledge required by the companies as described in Section 3, is covered within the core or elective courses before the internship takes place.

6 Conclusion

In this paper we described the design of a master's level curriculum in data science and artificial intelligence based on a gap analysis and assessment of needs from three case studies in Indonesia, Sri Lanka, and Thailand. On the one hand, we realize that in these countries many of the master's programmes we collected data from are generic master's in computer science and contain only few courses related to data science or artificial intelligence. On the other hand, we proceeded via a systematic process by gathering information from several companies in Indonesia, Sri Lanka, and Thailand to understand the needs and the gaps. Based on all information, we developed a novel curriculum for a master's programme in DS&AI that covers the main areas of the fields, it allows flexibility of study depending on the regional needs, and it comprises an internship for students to gain work experience. The approval and accreditation of the developed curriculum to be delivered by eight participating institutes in three countries are underway; hence, the programmes are expected to be launched in fall 2020 or spring 2021.

Acknowledgements

The authors would like to thank the European Union *Erasmus+* programme which provided funding through the *Capacity Building Higher Education* project on *Curriculum Development in Data Science and Artificial Intelligence*, registered under the reference number *599600-EPP-1-2018-1-TH-EPPKA2-CBHE-JP*.

7 References

Abduljabbar, Dia, Liyanage, and Bagloee (2019). Applications of Artificial Intelligence in Transport: An Overview. *Sustainability*, Vol. *11, No. 1,* pp. 189—212.

Bundy (2017). Preparing for the future of Artificial Intelligence. *AI & Society,* Vol 32, pp. 285—287.

Carriço. (2018). The EU and artificial intelligence: A human-centred perspective. *European View*, Vol. 17, No. 1, pp. 29–36.

Gomber, Koch, and Siering (2017). Digital Finance and FinTech: current research and future research directions. *Journal of Business Economics,* Vol 87, pp. 537–580.

Mathur, Aneja, Anwar, Shridhar, and Sanchez (2019). *Future of Work in Sri Lanka: Shaping technology transitions for a brighter future*, TANDEM research report, International Labour Organization, pp. 1—98.

Jiang, Jiang, Zhi, Dong, Li, Ma, Wang, Yilong, Dong, Shen, Haipeng and Wang (2017). Artificial intelligence in healthcare: past, present and future. *Stroke and Vascular Neurology,* Vol. 2, No. 4, pp. 230—243.

McKinsey Global Institute (2017a). *Artificial intelligence: Implications for China*.

McKinsey Global Institute (2017b). *Artificial intelligence: The Next Digital Frontiers?*

McKinsey Global Institute, (2017c). *Artificial intelligence and South-East Asia's Future.*

Quality Assurance Netherlands Universities (2015). *Artificial Intelligence Academic Programmes in the Netherlands: A State-of-the-Art Report.*

Skinner, Saunders, and Beresford (2004). Towards a shared understanding of skill shortages: differing perceptions of training and development needs. *Education + Training*, Vol. 46, No. 4, pp. 182—193.

Vairis, Loulakakis, and Petousis (2014). The role of internships in a higher education institute. In QScience Proceedings of the World Congress on Engineering Education 2013, pp. 27—34.

Wittayasin (2017). Education Challenges to Thailand 4.0. *International Journal of Integrated Education Development*, Vol.2, No.2, pp 29—35