Extended Abstract: Making AI work for skills-based training: A case study

R. Robson[1], E. Kelsey[1], A. Goel[2], L. Egerton[1], S. Nasir[1], M. Lisle[2], A. LaFleur[1], and E. Robson[1]

[1]Eduworks Corporation, USA. robby.robson@eduworks.com. [2]Georgia Institute of Technology, USA.

Abstract — Motivated by the need to upskill workers rapidly and equitably in response to changes in work and the workplace, a team funded by the US National Science Foundation Convergence Accelerator program has developed and piloted an application called *SkillSync* connects companies to college non-degree programs and facilitates the exchange of training requirements and relevant training opportunities. This abstract describes SkillSync and how Artificial Intelligence (AI) services are used in SkillSync to automate skills extraction and align training resources with required skills. This abstract then identifies issues we encountered in developing SkillSync and discusses lessons learned. The last section discusses further applications.

1 Introduction

The work presented here is supported by the US National Science Foundation (NSF) [1] and is motivated by the need to upskill millions of workers in response to changing job requirements. Since non-degree programs at local colleges are a natural but under-utilized source for training, we are developing an AI-enabled web app called *SkillSync* that connects companies to colleges for upskilling workers. SkillSync has undergone extensive design exercises with focus groups and been successfully used in three live trials, with additional pilots scheduled through 2022.

SkillSync takes a skills-based approach that requires the ability to identify and compare knowledge, skills, and abilities (KSAs) [2] in unstructured text. We have applied recent advances in natural language understanding (NLU) and machine learning (ML) to develop *AI services* that do this. These are exposed through application programmer interfaces (APIs) that are used by the SkillSync app and are available to other applications. This abstract describes the app, lists the AI services, identifies the issues we faced and overcame, and ends with takeaways and applications.

2 The SkillSync App

Our initial goal was to bridge the gap caused by the rapid pace at which industry needs evolve and the much slower pace at which academic programs operate and change. To understand the problem more deeply, we met with human resources (HR) and talent managers, directors of college continuing and professional education programs in the Atlanta area, the Business Higher Education Forum and the University Professional and Continuing Education Association. Our findings led to the development of an app that digitizes the connection between employers and colleges for the purpose of upskilling incumbent workers.

This app has two user types of users: *company users* responsible for HR, talent development and training at a company and *college users* responsible for coordinating (non-degree) training programs with company customers. Company users create *training requests* that identify skills to be acquired, together with information about the existing skills potential trainees possess. They then publish requests to specific colleges or to a marketplace. College users receive these requests, create *training proposals* that respond to them, and send them (via SkillSync) to company users for review and action [3].

Company users can search skills from many external sources (e.g., O*NET) by job title or keyword. They can prioritize skills, add new skills, and automatically extract skills from job descriptions they upload. In the Minimum Viable Product (MVP) release, users will be able to see skills trends derived from millions of live job postings from the National Labor Exchange (NLx) [5]. College users can import course data and search and select training offerings. When a college user selects a set of course offerings, SkillSync displays an *alignment score*, a number between 0 and 100 that indicates how well the selected offerings cover the skills in a training request. SkillSync also shows users how adding new offerings will affect the alignment score. An intelligent agent called *AskJill* provides help and answer questions about the app. The SkillSync web site includes up-to-date feature lists and to view videos that show SkillSync in action [6].

3 AI Services

SkillSync uses five AI services: *KSA Extraction* identifies Knowledge, Skills, and Abilities (KSAs and relationships among them in unstructured text; *KSA Generation* generates a prioritized list of the most relevant KSAs for a job title or description; *Alignment* computes an alignment score between a KSA and a course description by matching explicit and latent concepts, *CII Removal* replaces *Company Identifiable Information* (company names, brands, trademarks, technologies, processes, people and locations) with generic equivalents; and *AskJill* **i**nterprets and answers use questions about SkillSync in real time via a text interface. The first three are core to the operation of SkillSync and to skills-based talent management. The fourth is required to meet commercial requirements for accessing job data but is applicable to other use cases, such as scrubbing military personnel records. The fifth is an active area of research with the potential to increase the usability and trustworthiness of AI-enabled training and talent management applications.

All AI services use large, pre-trained language models such as BERT [7] and GPT-2 [8] as starting points, with transfer learning used to fine-tune pre-trained language models to perform specific tasks in specific domains. In CII Extraction, transfer learning was applied to an off-the-shelf pre-trained BERT model. The first three are based on language models pre-trained from scratch with data drawn from Wikipedia, a curated subset of the Common Crawl database of web crawl data, a database of job postings provided by the National Labor Exchange, and a dataset consisting of curated, open-source textbooks, course descriptions, and training materials.

AskJill is being developed at the Georgia Institute of Technology (Georgia Tech) by the Design & Intelligence Laboratory at [9]. It is based on technology that analyzes and answers student questions about courses based on the content of syllabi [10] and that has been used to explain the design and operation of an interactive tutoring system [11]. AskJill uses a two-dimensional hybrid ML and semantic processing model. An NLP-based model is trained to classify the *intent* of a question. A semantic processing layer converts intent into a structured query, searches data in structured and unstructured knowledge bases, and formulates a natural-sounding response.

4 Issues Encountered

Several issues were encountered (and overcome) in developing AI services for SkillSync. The first was bias. Language models can reflect gender, ethnic, racial and other biases inherent in the sources used to train them [12]. This has been observed in job descriptions [13]. In response, we used multiple methods to reduce bias, as measured by how closely job related terms are to racial or gendered terms in embeddings, whether racial or gendered terms co-occur with an occupation, and whether there are associations between race or gender, an occupation, and positive or negative sentiment. Our bias reduction methods have reduced these measures, which is encouraging.

A second issue is data acquisition. Our language models require labelled training data for each occupational domain. Subject matter expertise is needed to label the data, and data labelling is time consuming and repetitive. This makes it hard to find appropriate labor. After trying alternatives, we turned to commercial data labelling firms. These were more costly and had longer turnaround times than anticipated, and the results often required re-working. A related issue was that most existing skills frameworks are in non-machine-actionable formats (PDF, Word™, or HTML) and the skills in these frameworks are often too high level or too contextualized to be useful. We spent considerable effort finding and curating skills frameworks and converting them into machine-actionable data.

The time required for AI service development was also an issue. ML pipelines can configure and automate repeatable execution of model and generation [14], but the tools that these are complex and still evolving. In practice, they need specialized knowledge to operate and can be fragile. Significant collaboration among data scientists, data engineers, and IT is needed to configure, optimize and debug ML pipelines and the cloud-based computational environments used for model training. The computations themselves can take days and often cannot be parallelized. This slowed the development of AI services.

The alignment score was another issue. We wanted this score to reflect the match between prioritized set of KSAs and a set of course materials, but we knew of no theory that rigorously defined this match. As a result, we took an empirical approach in which success was gauged by user acceptance and whether the score behaved as expected when course materials were added or removed. The score we used in trials passed the “sniff test” in that it increased and decreased in a logical fashion and the results made sense to users, but more work is needed to ground the alignment in theory.

Another issue arose from a contractual requirement. SkillSync analyzes KSAs in job postings from the NLx to detect skills trends. Use of these postings is governed by an agreement that requires removing CII from job postings and KSAs to avoid revealing competitive information. This is a challenging NLP problem that we solved with algorithms that are layered on a version of named entity recognition (NER) [15] that incorporates *attention* [16].

The last issue concerns AskJill, which was first designed to answer questions about course materials. In focus groups, we discovered that company were interested in seeing the actual course descriptions, for which we did not need AskJill. AskJill is now designed to answer questions about the app, its operation, and its algorithms with the goal of improving transparency, explainability, and usability [17] (as well as providing contextual help). This shift requires more sophisticated knowledge-based reasoning, which is being developed using Task-Method-Knowledge (TMK) models that Georgia Tech has used for meta-reasoning in a variety of AI Systems [18] [19].

5 Applications and Lessons Learned

SkillSync and its AI services can support a broad range of applications. The techniques used for CII removal can be used to redact sensitive documents, we are exploring how to use the app to help military and government personnel find voluntary education, and KSAs collected from NLx will be used to identify skills trends and in-demand skills. This information will be provided to company and college users and disseminated in the NLx research hub [20]. AskJill is intended to be a generalizable asset that can be trained to provide help and increase trust in other apps.

In addition to producing generalizable technology, our work has taught us two important lessons that apply to anyone who wishes to develop AI-enabled skills-based training and talent management applications. The first is that while AI may be supported by ML packages, pipeline tools, and pre-trained models, there are factors (such as data acquisition) that can lead to unanticipated costs, and it is not unusual to encounter non-standard challenging requirements (such as bias reduction and CII removal in our case). The need for in-depth knowledge of ML, NLU, and related data science and engineering processes should not be underestimated. The second lesson is that apps like SkillSync operate in sociotechnical environments where technology and human behavior are intertwined. We made numerous pivots in the design of the app and its underlying AI services based on focus groups and trials and real-world business requirements, and we continue to fine tune many aspects as we engage with more and more diverse users. Underestimating the need for end-user input is a more fatal error than underestimating the complexity of the problems faced in transitioning to skills-based talent management and in using AI to support this transition.

6 References

[1] NSF, *Award #2033578,* 2020.

[2] US Veterans Administration, “What Are KSAs? - VA JOBS,” [Online]. Available: http://www.va.gov/jobs/hiring/apply/ksa.asp. [Accessed 27 December 2021].

[3] BHEF, “The Business Higher Education Forum,” [Online]. Available: https://www.bhef.com/. [Accessed 27 December 2021].

[4] UPCEA, “About UPCEA,” [Online]. Available: https://upcea.edu/about/. [Accessed 27 December 2021].

[5] NLx, “National Labor Exchange,” [Online]. Available: https://usnlx.com/. [Accessed 27 December 2021].

[6] Eduworks Corporation, “SkillSync,” [Online]. Available: https://www.skillsync.com/. [Accessed 27 December 2021].

[7] J. Devlin, M.-W. Shang, K. Lee and K. Toutanova, *Bert: Pre-training of deep bidirectional transformers for language understanding,* Google, 2019.

[8] A. Radford, J. Wu, R. Child, D. Luan, D. Amodel and I. Sutskever, “Language models are unsupervised multitask learners,” 2019. [Online]. Available: https://cdn.openai.com/better-language-models/language_models_are_unsupervised_multitask_learners.pdf. [Accessed 28 December 2021].

[9] Georgia Tech, “Design & Intelligence Lab,” [Online]. Available: https://dilab.gatech.edu/. [Accessed 08 January 2022].

[10] A. Goel and L. Polepeddi, “Jill Watson, A veritual teaching assistant for online education.,” in *Education at Scale: Engineering Online Learning and Teaching.*, C. Dede, J. Richards and B. Saxberg, Eds., Routledge, 2018.

[11] A. Goel, V. Nandan, E. Gregori, S. An and S. Rugaber, “Explanation as Question Answering based on Unser Guides,” in *AAAI-2022 Workshop on Explanation in Agency*, 2022.

[12] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama and A. T. Kalai, “Man is to computer programmer as woman is to homemaker? debiasing word embeddings,” *Advances in neural information processing systems,* vol. 29, pp. 4349-4357, 2016.

[13] H. R. Kirk, F. Volpin, E. Benussi, F. Dreyer, A. Shtedritski and Y. Asano, “Bias Out-of-the-Box: An Empirical Analysis of Intersectional Occupational Biases in Popular Generative Language Models,” *Advances in Neural Information Processing Systems (NeurIPS 2021),* vol. 34, 2021.

[14] A. Barrak, E. E. Eghan and B. Adams, “On the Co-evolution of ML Pipelines and Source Code-Empirical Study of DVC Projects,” in *2021 IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER)*, Virtual, IEEE International Conference on Software Analysis, Evolution and Reengineering (SANER).

[15] D. Nadeau and S. Sekine, “A survey of named entity recognition and classification,” *Lingvisticae Investigationes,* vol. 30, no. 1, pp. 3-26, 2007.

[16] A. Vaswani, N. Shazeer, N. Parmar, J. Uskoreit, L. Jones, N. A. Gomez, L. Kaiser and I. Polosukhin, “Attention is all you need,” in *31st Conference on Neural Information Processing Systems (NIPS)*, Long Beach, CA, 2017.

[17] K. Wiggers, “What is explainable AI? Building trust in AI models,” VentureBeat, 26 November 2021. [Online]. Available: https://venturebeat.com/2021/11/26/what-is-explainable-ai-building-trust-in-ai-models/. [Accessed 27 December 2021].

[18] W. Murdock and A. Goel, “Meta-Case-Based reasoning: Self-Understanding for Self-Improvement.,” *Journal of Experimental and Theoretical Artificial Intelligence,* vol. 20, no. 1, pp. 1-36, 2008.

[19] A. K. Goel and S. Rugaber, “GAIA: A CAD-Like Environment for Designing Game-Playing Agents,” *IEEE Intelligent Systems,* vol. 32, no. 3, pp. 60-67, 2017.

[20] NLx, “Welcome to the NLx Research Hub,” National Labor Exchange, [Online]. Available: https://nlxresearchhub.org/. [Accessed 29 December 2021].