Lessons Learned in Human-Artificial Intelligence Teaming in Business Processes

Michael J. Mendenhall [1*], Gilbert L. Peterson [1,2], Alexander Graves [1,3], Jonathan W. Butler [1,4]

[1]Air Force Research Laboratory, Autonomy Capability Team 3 (ACT3)
[2]Air Force Institute of Technology
[3]University of Dayton Research Institute
[4]Mile2
[*]michael.mendenhall.1@us.af.mil

Abstract

There are three teams that exist in research, development and deployment of a business process focused Artificial Intelligence (AI) system. These include the customer team, AI team, and User Experience/Interface (UX/I) team. This article presents six (6) lessons team members need to learn and adopt in order to be successful: the problem that needs to be solved is not always obvious; an AI business process automation system still needs people; State of the art and user expectations need to be aligned; Adding AI-automation to a business process is more than data in and data out; be cognizant of shifting workload; and the AI-automation may have to operate as a silent partner. These lessons are discussed within the context of the research and development of a Human Resource Apprentice to assist the evaluation of resumes against the specialized experience required for the advertised position.

Introduction

Developing and integrating Artificial Intelligence (AI)-enabled systems into complex business processes requires a deliberate co-creation process consisting of team members of varied backgrounds and experiences; AI scientists, engineers, and developers; and human factors scientists and engineers. As in Figure 1, the teams must work together and frequent detailed communications (indicated by arrows in Figure 1) are essential in order to determine the correct problem, create suitable design, and integrate the system properly into the customer's business processes.

This article describes the development of a *Human Resource Apprentice* (HRA), an AI automation that aides staffing specialists in assessing applicants against position requirements where that software must be integrated into existing hiring processes. This paper highlights six (6) lessons learned during the development of the HRA prototype. In describing the problem and the technical approach, the lessons are described and the communications between the Customer and AI teams, the Customer and UX/I teams, and the AI and UX/I teams are highlighted.

The lessons are presented in context of when they arise during the research and development process. Often the lessons appear at more than one time due to the types of communications occurring between the different groups at that phase of the development.

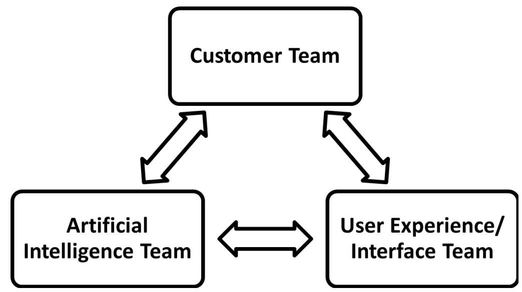

Figure 1: Communications (bi-directional arrows) between customer, AI, and UX communities are critical when developing artificial intelligence (AI)-enabled systems to support complex business processes.

The Problem: Hiring Civil Servants

Hiring Civil Servants in the United States Government (USG) is a complex and time-consuming business process. Position advertisement, applicant submission, and management of the hiring process is centralized through software solutions owned and hosted by the Office of Personnel Management (OPM). USG Departments and Agencies have their own staffing specialists that manage their interests using OPMs infrastructure. In Fiscal Year (FY) 2021, there were nearly 20M applicants for over 350K positions across the USG where the Department of the Air Force (DAF) received on the order of 1.5M applications. Given the complexity of the business process and the relative size of the number of positions and applicants to the number of staffing specialists, statutory hiring time-lines are rarely met.

Multiple customer and technical team communications were required to identify where automation could best assist staffing specialists. Figure 2 presents the process. Purple (Step 1 and Step 2) are pre-advertisement tasks. Orange (Step 3 and Step 5) are applicant gathering and prescreening tasks. Light blue (Step 4) are applicant tasks. Black (Step 5a and Step 6a) are non-referral notification tasks, where a non-referral means the individual will not go forward to the hiring manager and is no longer considered for the position.

Green (Step 6, Step 6b, and Step 7) are qualification and referred candidate related tasks, were a referred candidate is forwarded to the hiring manager and may be offered an interview in Step 7. Step 6 has the highest impact on staffing specialist workload and offers the best return on investment.

Lesson 1: The problem that needs to be solved is not always obvious.

In the initial problem exploration, there was no real sense of complexity associated with the problem beyond a basic matching automation against structured or semi-structured data. There are existing solutions and products that could be easily adapted to address customer needs. However, it was learned that Step 6 marked heavy staffing specialist involvement where they perform a detailed review of the resume and other submitted materials in response to **all** hiring requirements in order to make a referral determination.

But which of the multiple tasks should be addressed first? The customer internally reports measures of performance (MoP) as part of their own self-evaluation and business process improvement goals. Their MoPs pointed to eligibility verification task in Step 6, which is evaluating the candidate's resume against the specialized experience statement (SES) as the most time consuming part of the evaluation process.

An independent cognitive task analysis performed by an industrial psychologist was performed to make sure that the SES evaluation was not a symptom and that the underlying issue was elsewhere in the process. For example, internal organization workflow management issues, or software or hardware infrastructure challenges. The psychologist verified that evaluation of the resume against the SES is the cognitively taxing task and the most time consuming.

Upon further review, it was learned that SES's are not consistently structured, contain several complex statements, and lack the entirety of the content required to perform an evaluation (e.g., statements such as "must have (1) year of specialized experience at the next lower grade"). Additionally, resume layouts vary greatly as does the quality of their writing. From the perspective of training data, it turns out that the content in SES's and resumes alone are not sufficient to create an acceptable knowledge base.

System Overview

Figure 3 shows a diagram of primary AI components used to process resumes and evaluate them against the SES. Applicant resumes submitted in document form are segmented according to an OPM ontology. After resumes are segmented, they are evaluated against individual requirements in the SES and a score that indicates match quality between each SES statement and each resume sentence is maintained. Content is presented to the staffing specialists so that they may interact with them and create notes that track SES-resume matches to be used in their later forwarding steps.

Lesson 2: An AI business process automation system still needs people.

Customer team members were concerned about AI-enabled systems taking their jobs. Though this is something reported in the literature, it was a new encounter to the degree witnessed. As shown in Figure 3, staffing specialists are an integral part of the automation.

Artificial Intelligence

Lesson 3: State of the art and user expectations need to be aligned.

"We have doubts in the ability of the AI to make sense of specialized experience statements (SES)." "You will not be able to make a program that can out perform me." These are legitimate issues presented by the customer team. Casting the AI-automation as an apprentice to the staffing specialist set the right expectations. The "Human Resource Apprentice" can be expected to work with and for the staffing specialist, learn from ongoing interactions, and provide efficiencies.

Deliberate communications between the AI and the customer teams eased their minds by giving them the right understanding of the capabilities and limitations of AI for natural language processing and understanding problems. Similarly, deliberate communications on the same topic gave the UX/I team insights into how they should approach the human-machine teaming aspects of the problem.

From the AI development perspective, all that is required is that SES's and resumes are inputs and the output is a tuple that contains a similarity score for every requirements-sentence pairing between the SES and resume. Users can easily select the matches that make the most sense through a simple user interface.

Resume Segmentation

Lesson 4: Adding AI-automation to a business process is more than data in and data out.

Often, in adding an AI to a business process a majority of work becomes interfacing with existing data stores and user interfaces. There has been a lot of success when the data is already well structured and complete and when the problem is well defined. For data science heavy projects, the CRoss Industry Standard Practice for Data Mining (CRISP-DM) (Chapman et al. 2000) may suffice, or one of several extensions developed to make it more extensible (Martínez-Plumed et al. 2021). Our experience with this domain and other NLP/NLU domains are that often the data available for the task leads to needing to create more than one AI capability to provide a customer solution. The work in (Martínez-Plumed et al. 2021) offers a dynamic model that uses CRISP-DM at its core, where the dynamic nature is represented as different trajectories through several different process elements. Even this view required a heavy lift to address the complexities found in AI automation with multiple, connected, moving pieces that must be teamed with the people executing the business process.

Resumes have a conceptual structure, but each differs in layout, format, heading terminology and content style. Before creating an approach to evaluate resumes against the SES, resumes need to be segmented. Resume segmentation occurs for four reasons. First is to focus the evaluation of the resume against the SES in order to reduce superfluous findings. Second is to improve traceability between

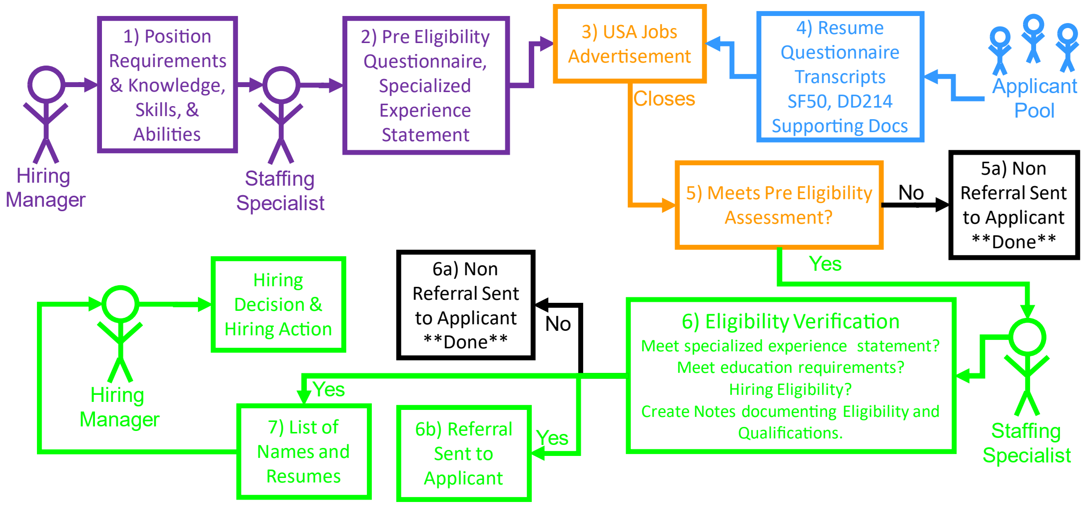

Figure 2: The process for DAF civil servant hiring.

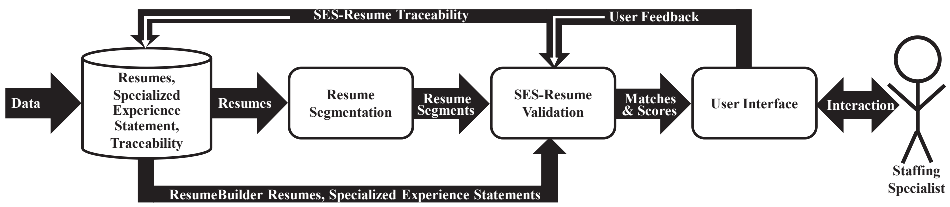

Figure 3: Overview of the HRA developed to support DAF civil servant hiring.

SES statements and passages in the resume. Third is to improve the experience during staffing specialist-HRA interaction. Fourth is to posture the system for evaluation tasks beyond the SES (*e.g.*, education and professional certification requirements).

Applicants can submit a resume as a document or use the built-in resume builder application in USAJobs. The USAJobs resume builder application stores content in a relational database, which provides semantically meaningful labels for resume segments. Resumes are segmented and placed in an intermediate form as structured data maintained as a <key,value> pair.

For those resumes submitted as documents, a resume segmentation system (Figure 4) was developed. Documents first go through *document layout analysis* where individual sections of text are identified. The identified sections are sorted and sent to *transcription* where optical character recognition (OCR) is performed. Lastly, the extracted OCR content is sent to *labeling* to categorize each identified region of text based on its content (*e.g.,* work history, education, and not applicable).

Implementation Details

Document segmentation is largely a machine vision problem. As such, all documents are converted to images where semantic segmentation models such as Detectron2 (Wu et al. 2019) can be fine-tuned to segment the document as defined by the ontology.

The semantic segmentation system is a series of individual micro services with web application programming interfaces (APIs) to handle data from user requests and/or other micro services or interfaces. Each micro service is designed to handle standard/common data formats and perform one function.

The micro service architecture offers a flexible implementation to address multiple document segmentation tasks in the HRA. In terms of resumes, the segmentation identifies areas that include prior position titles, descriptions and dates as well as educational elements such as degree, institution, and year of conferral. Because of the modular design, it is possible to repurpose the system to also parse transcripts and other forms used by the staffing specialist.

Segmentation Results

One advantage of the individual microservice architecture is that evaluation can be done in two dimensions: evaluating each service independently and an end-to-end evaluation. The data used in the evaluation was anonymized content from ResumeBuilder populated into custom resume templates. The resume templates allow easily-obtainable truth data for the layout analysis services and the anonymized content allows for truth data for the transcription services. The resume templates used in this evaluation were not seen by the layout analysis models during model training in an effort to remove any biases from the model. For the evaluation, there were three (3) resume templates used to format

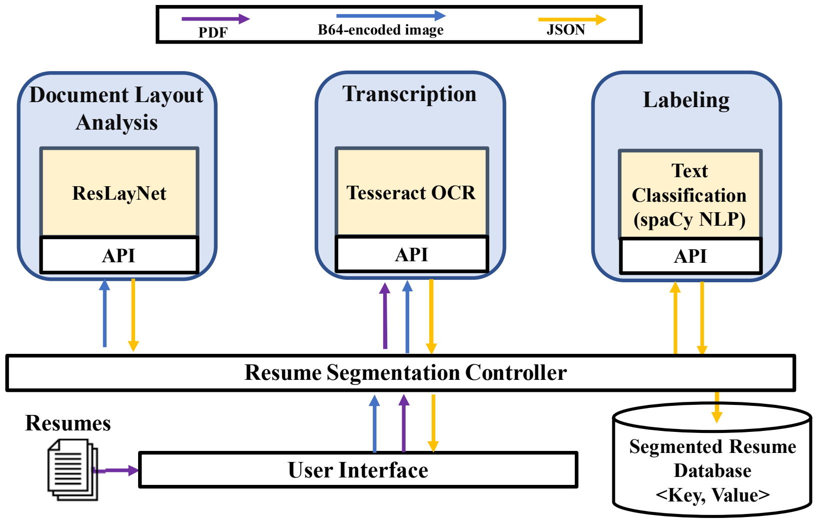

Figure 4: The semantic segmentation architecture consisting of three stages: layout analysis, transcription and labeling.

Table 1: Evaluation for the resume layout microservice.

Class	Precision	Recall	F1	#Images
Body	0.87	1.00	0.93	3,047
Header	1.00	0.74	0.85	1,685

Table 2: Evaluation results for the transcription microservice.

# Words	# Characters	WER	CER
766,330	5,230,386	5.24%	2.41%

450 different resumes of varying length resulting in 1,800 pages of (labeled) resume content.

Table 1 provides a performance assessment of the resume layout analysis using the resume layout network (ResLayNet). ResLayNet uses Facebook AI Research's (FAIR) Detectron Mask R-CNN semantic segmentation model (Wu et al. 2019) fine-tuned to identify and segment section headers as well as blocks of text in resume documents. Though the current task is to identify experience and education sections, the evaluation for resume layout includes other resume sections as well (*e.g.*,publications, awards, certifications, etc).

Google's open source Tesseract is used for transcription since it has acceptable levels of performance out of the box, but also allows for custom-trained models for domain-specific character sets that the provided model might incorrectly recognize. The model used in the evaluation is Google's release model. The evaluation for transcription service tests both the word error rate (WER) and character error rate (CER). Table 2 summarizes the results.

The labeling microservice uses spACY for classifying extracted blocks of text as either experience or education. The classification is based on an evaluation of the text content and not the position in the resume. Table 3 summarizes the results.

Table 3: Classification results for the labeling microservice. Row entries are actual classes and column entries are predicted classes.

	Experience	Education	Missed	Total
Experience	437	0	13	450
	97.1%	0.0%	2.9%	
Education	105	331	14	450
	23.3%	73.6%	3.1%	

An end-to-end test is conducted using ResLayNet for layout analysis, Tesseract for transcription, and spACY for classifying each extracted block as either work experience, education, or not detected. As accomplished previously, resume layout results report text sections that are correctly segmented, under-segmented (the model misses part of the content from a section), and over-segmented (the model includes multiple section content as one section). The WER and CER are computed for text sections associated experience and education only are presented. Finally, experience and education segment classification results are reported in a form similar to a contingency table. Table 4 compiles the results of the end-to-end evaluation.

Specialized Experience Evaluation

Initial discussions with the customer team focused on matching position description with knowledge, skills and abilities from the SES. For example, a position asking for software engineering experience would need to be able to match an applicant's position description including common

Table 4: Results from the end-to-end evaluation using ResLayNet for layout analysis, Tesseract for transcription, and spACY for text classification.

	Total Segments	Correct Segments	Under Segmented	Over Segmented		Experience	Education	Missed	Total
Overall	4,732	2,642	1,526	454	Experience	740	0	74	814
WER	-	6.86%	-	-		90.91%	0.00%	9.09%	-
CER	-	3.13%	-	-	Education	54	206	11	271
	-	-	-	-		19.92%	76.02%	4.06%	-

software engineering terms, such as Agile, SCRUM, etc. A portion of discussion was associated with **Lesson 3** (State of the art and user expectations need to be aligned.) The AI team wanted to communicate to the customer team that there is a tradeoff between a general model for the broad problem scope and narrowly focused domain models for depth of knowledge accuracy.

The other prevalent lesson is **Lesson 4** (Adding AI-automation to a business process is more than data in and data out.) SES statements are fairly standard for a specific job series and for a specific grade. This means that there is a lot of consistency for a job/grade series. The per-position differences exist in the technical capability needs that tend to be appended at the end of the SES. This is the matching component. However, a second aspect to the problem is the sentence and logic complexity found in the SES.

Specific to the SES, the challenge is building models of concept relations and then applying the relations during evaluation. The concept relations include: compound phrases, anaphora (pronouns), booleans, lists, and ”musts” vs ”optionals”. Parsing of these currently leverages entity recogntion and link grammars to get an initial parse that is then manually segmented for use by the HRA (Goertzel, Pennachin, and Geisweiller 2014). Concept relation application requires a semantic match as well as business logic connections to related fields such as dates or degree titles.

When selecting a model, care must be taken to avoid bias against applicant populations as much as possible. The NumberBatch (Speer, Chin, and Havasi 2017) embedding (the duel to the ConceptNet knowledge graph) is used. It is selected as it uses algorithmic debiasing to remove known bias types (Speer 2017) and it has been evaluated for demographic fairness in (Sweeney and Najafian 2019) with good results. Matching return is a weighted function over the dot product, Euclidean, and cosine distance measures.

The output is a tuple that contains a resume sentence and its location in the resume, a SES sentence and its location in the SES, and a similarity score. This information is used by the UX team to present findings to the staffing specialist for evaluation.

User Experience (UX)

Staffing specialists were initially reluctant to share detailed insights into their process due to fears, doubts, and misnomers that machines would be stepping in to take over their work (**Lesson 2:** An AI business process automation system still needs people and **Lesson 3:** State of the art and user expectations need to be aligned). Communicating the human-machine teaming strategy early on was critical. Establishing the machine as an apprentice puts the machine in the right place for this effort (the machine works for the staffing specialist), but it lacked specificity. It was necessary to educate the customer on how the UX and the AI together could be used lighten the cognitive burden of scanning the entire resume for each requirement in the SES and how that would allow the staffing specialist to focus on more meaningful work sooner in their processing.

In order to reach the state in which staffing specialists are comfortable with the idea of a machine doing part of their workload, it is important to understand their way of thinking and how they complete their tasks. Second, building their trust is vital if it is expected of them to trust working with an AI automation system. The volume and complexity of their task makes it overwhelming and the goal is to maximize their efficiency and effectiveness by automating mundane tasks and redirecting their attention to more critical tasks that the machine cannot do (Woods 2018).

Discovery: How Staffing Specialists Work

Lesson 5: Be cognizant of shifting workload.

Since the HRA should learn from user interaction a simple deliberate feedback mechanism should suffice: provide a rank-ordered list of candidate matches and simply have the user provide a thumbs up for an acceptable match and a thumbs down for an unacceptable match. Thumbs up responses can be used to create justification notes and both thumbs up and thumbs down responses can be used for learning. This early design consideration and prototype over simplified the HRA experience the staffing specialist required based on the complexities of their task and the varying experience levels of their team members. The user feedback was pretty clear: this is a terrible idea. The design concept made the customer feel like they were babysitting the AI automation; it made them feel like they were spending their time assessing the machine; and they were rightfully concerned about increased workload.

Workload is a big factor in creating the right UX. Care must be taken to not add friction to the way customers perform their work so as not to inhibit their ability to complete it. One thing that became clear is that staffing specialists do not want actual or perceived extra steps in their processes as it is perceived as making their job harder. The following examples illustrate where staffing specialist’s had concerns that the HRA would slow their process down or make their job harder: the current state-of-the-art AI model performance; extra affordance clicks within the user interface;

providing explicit feedback to the HRA; or needing to double check if the HRA declared a job candidate as qualified or not qualified.

What is the right user experience based on staffing specialist feedback? It depends because there is more than one user, each with a different way of working. During the elicitation process, staffing specialists demonstrated the use of multi-monitor setups (Figure 5), various ancillary browser sessions and applications to locate and document supporting information, and using hard tools such as pen and paper as fail-safes should the existing system go down.

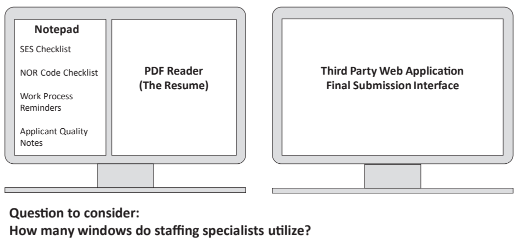

Figure 5: Staffing specialist monitor setup with multiple applications running in order to support the cognitive work.

The elicitation processes revealed the reality that staffing specialists have many approaches taking place in concert that are far more complex than what a new custom tool can be expected to address. That said, considering how to bring the staffing specialists complex world into a single capability to help streamline their workload is highly appealing to them. When doing so, care must be taken to ensure that baked-in machine bias and/or staffing specialist bias does not emerge.

Arriving at the Right UX

Lesson 6: The AI-automation may have to operate as a silent partner.

It became apparent that the AI automation needed to operate silently in the background, literally **and** in the subtle nuances of the user interface. The goal of the UX is to enable users to focus on identifying the unique problems they need to solve within the work, without adding more friction or fear to their already challenging process.

During knowledge elicitation sessions, a/b tests, and prototype demonstrations, the customer generally self-reported where their issues were in their current workflows and in the hypothesis presented to them. For example, two different staffing specialists mentally parse the SES into a handwritten checklist of sorts as an analysis aid and believe an automated solution for that process would be helpful. A third staffing specialist did not believe that senior staffing specialists would benefit from the automation concept, that it might be better suited for someone more junior. It was clear early on that it was necessary to demonstrate the value for junior and senior team members.

In order to support junior staffing specialists, the HRA would need an onboarding experience that demonstrates how using the AI automation will make them successful. Ideas such as a built-in acronym dictionary seemed useful to help build a lexicon within a large industry. In order to support a senior staffing specialist, streamlining the resume evaluation processes within a single well organized user interface (Figure 6) would be valuable. This would greatly reduce time spent on burdensome tasks allowing staffing specialists to apply more energy on making the final referral decisions, a task that is not well suited for the HRA due to legal and ethical concerns. A data-driven focus can be problematic as it may allow the staffing specialist to over rely on the HRA resulting in gross decision errors due to a lack of supporting data. However, a data-informed focus allows the person to make a rational and educated decision, amplifying the staffing specialist's ability to perform their tasks.

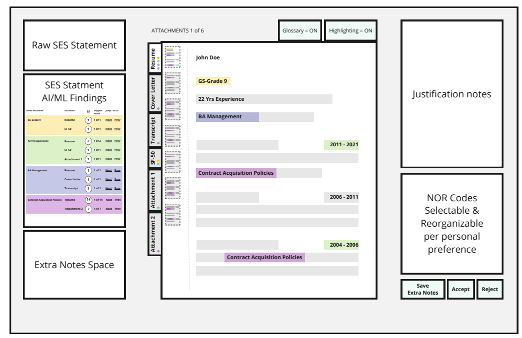

Figure 6: A basic low-fidelity wireframe that brings a staffing specialists common multi-windowed set of applications into a single user-interface.

The team explored various teaming strategies to include gamification mechanisms, loud strategies, and strategies that have the AI automation work silently in the background.

If the HRA highlights a string of text based on the AI automation's matching algorithm, it can only draw the staffing specialists attention to it in order for them to intervene and provide explicit feedback (the thumbs up/thumbs down example earlier). This is a loud action which as stated earlier makes the staffing specialist feel like they are babysitting the machine instead (**Lesson 5:** Be cognizant of shifting workload). The HRA uses a more silent approach by offering a plus (+) button which adds the machine highlighted text to the staffing specialists justification notes (**Lesson 6:** The AI automation may have to operate as a silent partner). This action also provides positive feedback to the AI automation from which it can learn.

One idea explored was to provide an extra notes section in the user interface that the AI automation could learn from behind the scenes. In the end, the thumbs up / thumbs down approach was left as an optional form of feedback. This allows the staffing specialist to provide strict feedback even if they did not use the content as part of the justification notes. The goal was to harness the staffing specialists time and energy as they do their normal work, while simultaneously addressing the information recall challenges faced by the AI automation in a frictionless and silent way (**Lesson 5:** Be cognizant of shifting workload and **Lesson 6:** The AI automation may have to operate as a silent partner).

The team also explored the use, disuse, misuse and abuse of automation to inform UX design. An important implication is to ensure biasing is not introduced by the staffing specialists or by the AI automation as a result of how the staffing specialist interacts with the system or how the AI automation results are presented. Failure of automation is a legitimate concern. Participants chose manual control if their confidence in their own ability to control the plant exceeded their trust of the automation, otherwise they chose the automation (Lee and Moray 1994; Parasuraman and Riley 1997). An area where this applies to the HRA is if the server hosting it were to go down. Can the staffing specialist's trust that the HRA will work offline as expected and when it comes back online their work remains intact? If yes, they will likely use the HRA. If no, they will likely continue using multiple windowed apps to perform their work should the system go down.

In designing the UX, care was take to avoid the abuse of HRA and the underlying AI automation. Abuse occurs due to design and implementation of automation solutions without considering the performance impacts of the person using it or where the person authority lies over the automation (Parasuraman and Riley 1997). It can also occur by inserting bias into the decision making process when decisions are made under uncertainty. Earlier discussions on forcing the user to provide a thumbs up/thumbs down is an example of where abuse could have been encountered. By making the thumbs up/thumbs down optional, and changing how the HRA obtains feedback, the first example of abuse should be largely avoided. For the second case, by not exposing the results of matching resume content with SES statements (Section) to the staffing specialist and by not rank ordering the results, the second case should be safely avoided.

After several iterations of the design cycle, both the customers and the AI team began to appreciate the flow of the interface. It was clear, intuitive and unobtrusive in that the staffing specialist could easily see everything that was pertinent to their work in one view. They understood that the AI automation was doing its best to surface relevant information to them in order to save time and allowing the staffing specialist to focus on the more important work of evaluating the candidates quality and credibility.

User Experience Design

The current HRA UX incorporates several of the aforementioned approaches and goals. Figure 7 address the need for centralizing the way staffing specialist's perform their work. The UX organizes functions in a left-to-right flow based on staffing specialist feedback. The as advertised SES is available on the top-left side of the UX and the AI automation's findings of the SES statements in the resumes are organized by short titles from the SES. The staffing specialist has the ability to turn highlighting on or off in various ways. For example, findings are highlighted in the document (middle of the UX) based on the scope of the short titles from the SES. Here, if one short title is expanded, those statements are highlighted. If all short titles are collapsed, all findings are highlighted in the document (middle of the UX). Another option is to show only custom highlights that the staffing specialist made, which also serves as an additional intervention to provide the HRA feedback. Additionally, custom highlighting is attributed to a the staffing specialist that performed it which supports various scenarios such as review or hand-off. If the staffing specialists selects the plus (+) with the snippet associated with the short title (bottom-left), the short title and the text highlighted in the document (middle) are added to the justification notes (top-right). The staffing specialist can also simply type notes in the text box (top right). At any point, the staffing specialist can save their progress by selecting *save* (bottom right). Additionally, by selecting the plus (+), the requirement from the SES and the highlighted text are used as feedback for the HRA. Once the staffing specialist completes the resume evaluation process, they can select the appropriate notice of referral (NOR) code (bottom-right) to the justification notes (top-right) and submit the completed package by selecting *submit* (bottom-right). The acronym glossary, as described earlier, was not implemented as it proved overly complex at this time.

What remains a hallmark for plausible solutions for the UX is that everything has a purpose and it is meaningful to the staffing specialist. The affordances must be clear and understandable. The work of the machine and staffing specialist feedback to the machine must remain unobtrusive. Most importantly, the overall teaming relationship between staffing specialist and machine should result in a reduction of burdensome energy and be free of bias so that they can focus on determining an applicant's eligibility.

Conclusion & Future Work

Creating AI automation solutions for complex business processes is a difficult task. The journey presented in this article revealed six (6) lessons summarized here. **Lesson 1:** The problem that needs to be solved is not always obvious. **Lesson 2:** An AI business process automation system still needs people. **Lesson 3:** State of the art and user expectations need to be aligned. **Lesson 4:** Adding AI-automation to a business process is more than data in and data out. **Lesson 5:** Be cognizant of shifting workload. **Lesson 6:** The AI-automation may have to operate as a silent partner.

For future work, additional segmentation models will be added to support other documents submitted for review. This leads to extending the matching portion of the HRA to assist with validating education requirements and hiring eligibility requirements. Additional work is ongoing to extend the matching portion of the HRA. This includes investigating how general and domain-specific models can be used in concert to improve performance and how external content can be acquired and incorporated in order to provide additional context to the evaluation. An example of the former is using a general model like NumberBatch with a model tuned to a specific career field (e.g, AI in computer science) together to find relevant content in a resume. An example of the latter is creating a knowledge representation of acronyms and their typical context to improve information recall or offer a hover over service in the UX that correctly defines the acronym. For the UX, scalability will be supported by enabling auto loading all applicant documents into the system instead of downloading them and opening them separately in an appropriate viewer (e.g., Adobe Acrobat). Interaction

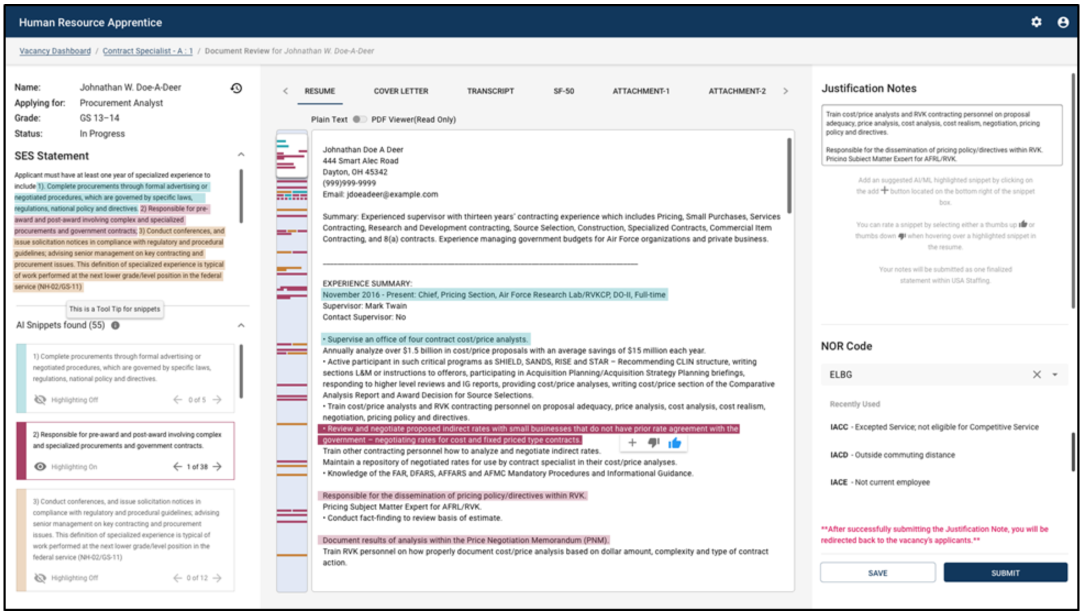

Figure 7: HRA resume review page.

with the other materials could be accomplished via a stepper control or using tab navigation. Tabs could be used to show content discovered by the AI automation through color encoded connections between applicant materials.

Acknowledgments

This work is funded by the Air Force Research Laboratory and the Air Force Material Command Personnel Directorate. The authors would like to thank the AFMC Personnel Directorate for their tireless support and participating. Without customer champion and involvement, the work described in this article would not be as impactful. We would like to thanks the Office of Personnel Management for their support and partnership as we both explore efficiencies in hiring through the use of AI-technologies. The authors would like to thank the ACT3 team members working on this effort: Josh Tang, Vahid Eyorokon, Elizabeth Hou, Andy Rice, Dr. James Okalica, Dr. Michael Vidulich, Dr. Lisa Douglas, Marisa Bigelow, Justin Dunham, Ren Estep, Jackie Scheid, Nero Lee, Caitlin Pennington, and Jennifer Davis.

The views expressed in this article are those of the author and do not necessarily reflect the official policy or position of the Air Force, the Department of Defense, or the U.S. Government.

References

Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; and Wirth, R. 2000. CRISP-DM 1.0 Step-by-step data mining guide. Technical report, The CRISP-DM consortium.

Goertzel, B.; Pennachin, C.; and Geisweiller, N. 2014. *Knowledge Representation Using the Atomspace*, 31–54. Paris: Atlantis Press. ISBN 978-94-6239-030-0.

Lee, J. D.; and Moray, N. 1994. Trust, self-confidence, and operators' adaptation to automation. *International Journal of Human-Computer Studies*, 40(1): 153–184.

Martínez-Plumed, F.; Contreras-Ochando, L.; Ferri, C.; Orallo, J. H.; Kull, M.; Lachiche, N.; Quintana, M. J. R.; and Flach, P. A. 2021. CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33: 3048–3061.

Parasuraman, R.; and Riley, V. A. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of Human Factors and Ergonomics Society*, 39: 230 – 253.

Speer, R. 2017. ConceptNet Numberbatch 17.04: Better, Less-Stereotyped Word Vectors. https://blog.conceptnet.io/2017/04/24/conceptnet-numberbatch-17-04-better-less-stereotyped-word-vectors/.

Speer, R.; Chin, J.; and Havasi, C. 2017. ConceptNet 5.5: An Open Multilingual Graph of General Knowledge. In Singh, S. P.; and Markovitch, S., eds., *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, 4444–4451. AAAI Press.

Sweeney, C.; and Najafian, M. 2019. A Transparent Framework for Evaluating Unintended Demographic Bias in Word Embeddings. In *ACL*.

Woods, L. 2018. Design's true impact and creating for 2 billion people. https://www.highresolution.design/new-gallery/jlxtpsnlgjv8s3knaa39u5qmm83z3e.

Wu, Y.; Kirillov, A.; Massa, F.; Lo, W.-Y.; and Girshick, R. 2019. Detectron2. https://github.com/facebookresearch/detectron2.