Review of: "SECURE II: Unlocking the Potential of Artificial Intelligence for Entrepreneurial Success"

Michael Falta[1]

1 University of Canterbury

Review "SECURE II: Unlocking the Potential of Artificial Intelligence for Entrepreneurial Success

P2: 'BM' – define abbreviation. Business model?

P2, 3rd paragraph: '*the*' – remove italics.

P2, last paragraph: ".. it [AI] can carry out … even impossible tasks…" – word choice. If a task is **impossible**, nothing can carry it out, not even AI.

P3, top: add a blank after '(1965)'.

P3, 3rd paragraph: you don't need to re-define "..Artificial intelligence (AI)…". Same applies to the next paragraph on the same page; and later in the last paragraph again; and on P4 again. I won't repeat this any longer: search the entire document so that repetitions of this sort are eliminated.

P3, 3rd paragraph: "…it [AI] has developed in an unstructured manner in the academic and professional literature.." is a false statement. AI was/is being developed by programmers and developers using computer languages, and not in the 'literature'.

P3, 2nd last paragraph: rewrite the sentence starting with "According to Corea …": it doesn't make sense.

P3, 2nd last paragraph: what is the difference between 'real-time' data analytics' and 'AI'. Considering the definition for AI you have cited on p2 (Jha & Topol, Schmidt) there is no difference!

P3 Heading "Artificial … techniques" are often classified into subsets including ML, NLP, robotics and machine vision. Change the heading to ML – AI techniques, because you review ANNs (which btw, is only 1 class of ML algorithms, so your review is incomplete. What about trees, forests, GMM, Bayesian classification, SME, …? So better include ANNs into the heading.

P3, last paragraph: I wouldn't rely much on 'research' which uses and associates words like 'proven links between explanatory and criterion variables'. There are few disciplines which prove relationships (theoretical subjects like mathematics, physics). In all applied fields, social sciences in particular, no theory ever proves relationships. We gather evidence in support of a phenomenon, we infer from our results that criterion and explanatory variables correlate or

associate. In fact, Popper's falsification idea is how research must be done – we are trying to disprove theories – and if we cannot, then the theory becomes stronger in the sense that more observations support it. Anything else is pseudo-science.

P4, 1st paragraph. Comment: I'm now 3 pages into the article and slowly but surely the reading becomes repetitive and vague. I haven't read any explicit statement or example which would specify which algorithm or firm software the authors mean to be AI? It's all about benefits here and there, but there is no substantive text. Let's go back to the AI definition used:

> 2353). Further, Schmidt et al. (2020), focusing on AI capabilities, defined AI as *the ability of organizations to use data, methods, processes, and people to create new possibilities for automation, decision-making, and collaboration that would not be possible by conventional means*".

I wonder what the authors mean to be AI? Furthermore, this definition states 'AI using people' which needs further elaboration. Also, AI using data, methods and processes = analytics (as mentioned above). Output being automation, decision-making support, again this is analytics. And the last bit of the definition 'would not be possible by conventional means'. Well, that depends on the capability of the person who performs the analytics – exactly equally would be the quality of AI depending on the capabilities of the person(s) who implement(s) the algorithm.

P4, 2nd paragraph: the authors say that AI=ML and DL techniques. No, there is an important difference: AI is computer software that mimics the ways that humans think in order to perform complex tasks, such as analyzing, reasoning, and learning. ML is a subset of AI that uses algorithms trained on data to produce models that can perform such complex tasks. Check also https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/artificial-intelligence-vs-machine-learning/#introduction

P5, 2nd paragraph: typo "Dep Learning models". When you use DL for deep learning, apply it through the text.

P5, "Neuro Symbolic AI": while this is fascinating research, how does it relate to the what the article means to explore? How does (NS)AI aid ex-ante decision-making for entrepreneurs? Really, the same applies to the entire p2/3: a literature review should clearly demonstrate a relationship with the intended research question. Currently, I am reading an overview of ANNs which I have read at least 50 times in other journal articles. I haven't learnt anything new. Where's the connection to an entrepreneur's tasks? In the introduction, I'd expect an explicit discussion on SECURE I and associated literature (if it exists), followed by a critique of SECURE I which would then motivate and justify SECURE II. I'm on page 5 now, and still haven't read anything about entrepreneurs (which industry sector?), their decision-making contexts and how the role of SECURE I or II (would) play out.

P6 finally attempts the link entrepreneur – business analytics. The 3rd paragraph is useful, but the 'two observations' need further exploring. Isn't the goal to eventually demonstrate which explanatory and criterion variables have been identified? List some examples from the GEM, VES and PSED. Unfortunately, the 2 paragraphs which follow, revert into vagueness: ex-ante measures, improving BM evaluation, detect patterns, improving BM parameters…

P6 would also benefit from clearly distinguishing 'novice', 'budding' and 'nascent' entrepreneurs. Are these the same, if yes, why confusing the reader, if not: explain.

P7: RQ (research questions) 2 and 3 are the same, you just replaced ANN with AI, used 'reduce' as opposed to 'eliminate' data uncertainty (where reduce is clearly the better word), and the objectives are also the same.

P7 contributions: finally, the reader reads about SECURE II. (Still nothing about SECURE I, though). Those 2 paragraphs should be on the first page of the introduction, and some more discussion about what 'data uncertainty' for entrepreneurs means. And why would human cognition on images and words have anything to do with entrepreneurs? What is your model for entrepreneurial decision-making?

P8, 1st paragraph: You list 3 features of data science, however, the 2nd and 3rd are the same 'reproducibility of social research'.

P8: The heading is 'the role of data science'. This has repetitive elements which are unnecessary. The main body uses the same text as above, but replacing AI with data science. The last 2 paragraphs are about research method – so the heading is inappropriate – and need much more discussion. Who beta-tested SECUREII? There is a lot of basic research elements missing. Descriptive statistics, dates, who did what, and how… Furthermore, what is the reader to make of the list of variables? Further, vague statements such as "The Ease of Doing … scale was applied" rather than explicitly stating what that scale actually is. Further, "collection of data was processed to devise Machine Learning models…" needs rephrasing, and an explanation what ML model the authors refer to? The ones from SECUREII, ANNs?

P8, "ML Modeling" starts with a 1st paragraph which continues to list 'advantages' of ML. That's unnecessary repetition. Get to the point of your research – we are now in the methodology section, so tell the reader what you've done.

P9, 1st paragraph – So 8 pages of ML, AI, ANNs, and now you are saying that you run regressions on a dependent variable 'Entrepreneurial capability'? But in the same paragraph you mention ANNs. Very confusing. Furthermore, the 2nd half of that paragraph is back at useless repetitions.

P9: "Performance statistics may have a greater variance"? and "Parameter estimates may have a more significant variance"? ??? How did you prepare the data set? Which dataset? And what is the reader to do with those two 'considerations'??

P9, 2nd paragraph: The list of analysis elements without any logical order and lack of justification has no merit. Which model did you evaluate? How does it relate to your research questions on P7? Then, the arbitrary soup of analysis statistics rather suggests the 'let's calculate all our analysis software can do, and let the reader find out if any is relevant'-type of research. Go slower: there is merit in understanding what you are doing. Explain why you use Kolmogorov-Smirnov, why you use Kendall's Tau and all the other dozen tests.

P9/10 "Symbolic AI". Is the reader to understand that symbols which represent construction permits, electricity connections etc are not recognized by entrepreneurs ex-ante, but would need AI to be detected as input data to predict

entrepreneurial success? If that's the case, it needs to be written as such and further expanded with a good argument why the human would be worse than AI in identifying symbols and associated meaning. If 'not', what's the point of this section?

P10/11: The presentation of the model in Tables 1 and 2: So in table 1 you present a set of widely used accounting measures (they are not complex as indicated in the table's heading) – fine. You also intend to use 'soft' measures, or non-financial information – great. How did you measure those metrics, i.e., how did you translate the words in the ordinal scale to numbers? Furthermore, the model shown in Figure 1 is complex. Firstly, let me just check: so you actually didn't develop any model – you are copying someone else's work? That's not so good. In any case, you need to explain Figure 1.

P11/12: Over so many pages you talk about ML, ANN, and then you run a logistic regression. That's a surprise, to say the least. On P12, you refer to Figure 1 claiming that it "predicts 'Entrepreneurial Capability'". I don't see how? These words don't appear in the figure. Same applies to 'Entrepreneurial Capability'. An equation would be good, so that the reader knows what you've done. Also, it says that the figure is credited to Mishra & Tripathi, however their publication doesn't contain this figure. You need to clearly display if the figure was reused, or if you developed it yourselves but included information reported elsewhere. Same applies to Figure 2 (P13). Where does this figure come from? Where does the data come from? Following on from this, the results in Figure 3 are meaningless, void of any possibility for contextual interpretation. Same applies to the rest of the analysis of Part 1.

P16/17, Figure 6. Figures need axis labels. Furthermore, the drawback of not explaining properly what SECURE1 and 2 are in the introduction results in that I have no idea what Figure 6 is representing. Somehow you managed to also produce some results using decision trees, which is explained in 6 sentences on P18. This is obviously not insufficient.

Part 3, p19ff, Ensemble model, I won't even review. This is the first time you mention this in the entire paper. There is no literature review, there is no model presented, there are no explanations how you have translated NSAI elements through EAS into ensemble modelling.

Some comments on the conclusion sections: P23 you conclude that including human knowledge into ML is beneficial. But at the outset of the paper, you claimed that AI would do a better job than people could ever hope for. I hope you see the irony in your conclusion?

The discussion section on P24 returns to vague statements which do not relate to the explicit results.

To make this an acceptable research contribution:

Discuss logistic regression, ANN and Ensembles in the introduction. Clearly state these are your methods. Show that they've been applied for x, but you are investigating y, hence your contribution. If you want to label ANNs and Ensembles as 'AI', fine, but misleading. In the lit review, report on articles which empirically demonstrate which independent (input) variables have been empirically demonstrated to have a relationship to ex-ante entrepreneurial decision-making. Clearly elaborate a model and discuss it adequately. Present your instrument. Present the responses, provide descriptive statistics. Do a preliminary data analysis and report findings. Only then proceed to the estimation/modelling part using

your 3 main methods to derive associations between dependent and input variables. Definitively, explain SECURE 1 and 2 properly. Furthermore, it seems to me that you are dealing with latent variables (e.g. entrepreneurial capability), so use methods which can handle such models (i.d. not logistic regression)