Encoded Bias

A reason to exercise caution in the embrace of artificial intelligence in the business realm

Grace Hind

Abstract:

Current trends suggest that artificial intelligence will play a significant role in the future of work. Due to the technology's efficiency and capacity for value creation, businesses are increasingly investing into artificial intelligence. However, this paper explains why they should exercise caution in this endeavour: Theoretical and empirical research from the academic realm suggests that artificial intelligence is marked by bias relating to gender, race, and other social factors. If it wishes to protect diversity and equality, whilst further integrating artificial intelligence, the business realm must address this issue.

Introduction:

Artificial Intelligence (AI) is essentially 'the ability of a machine to perform cognitive functions that we associate with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem solving, decision-making, and even demonstrating creativity'. (Rai et al., 2019). Although much of the public discourse surrounding AI in business has 'tended towards the dystopian' (Stephen, 2022), AI has already permeated a diverse range of areas in the business realm (from customer service, to analytical work) and most people are generally accepting of this. For instance, the majority of workers will see little significance in using a virtual assistant such as Siri or Alexa to retrieve information or manage their time. Furthermore, the intelligence augmentation approach seems to alleviate fears that AI will render human intelligence inutile.

Although businesses have seen many positive returns on their investments into AI, this piece will explicate a strong reason for the exercise of caution in the embrace of AI in the business realm – namely that machine-learning algorithms reflect human bias (West et al., 2019). This is a very pressing concern, as major companies have started to use AI in areas, such as the hiring process, where bias is of great consequence. However, the integration of AI into the business realm need not be halted completely as there is a way to mitigate the impacts of the problem of encoded bias; AI audit procedures such as capAI. (Floridi et al., 2022)

The rise of Artificial Intelligence in the business realm:

In the words of Andrew Stephen, Director of the Oxford Future of Marketing Initiative, 'over the past decade, the corporate world has been seemingly embroiled in a Gadarene rush to embrace AI, whether it's big tech racing to get AI products to market or companies hedging bets that pandemic-proof, always-on chatbots and other machine-learning systems can cut costs and increase productivity.' (Stephen, 2022) Current evidence suggests that this specific form of digitalization is proving beneficial for businesses; improving operational efficiency and value creation, (Zhou et al., 2021) so it is reasonable to assume that the trend will continue.

However, since the time when AI was merely science-fiction, rather than realised-fact, its development has been tied to the worry of machine dominance – a fear that humans will be outcompeted and even replaced by such technologies. As the capabilities of AI have evolved, this concern has remained a dominant theme in the academic literature surrounding digitalization, and it could be argued that businesses should take heed. This is because there is a risk that the increasing adoption of AI in the workplace will have a negative impact on employees' identifications with their jobs, and feed into individuals' fears of being replaced (Mirbabaie et al., 2022). These potential impacts could have the knock-on effect of creating disaffection and a consequent lack of productivity and company loyalty within the workforce.

Yet, this piece will demonstrate that the worry of machine dominance has perhaps been overstated. Whilst it is true that the rise of AI will shift the balance in businesses' investments into human versus physical/mechanical capital towards the latter (for instance investing into delivery robots, such as Amazon's "Scout", rather than employing more delivery drivers), it need not be true that the embrace of AI in the business realm will disadvantage the majority of employees.

This is due to the emerging trend towards intelligence augmentation (IA). The concept of intelligence augmentation is based upon the argument that artificial intelligence systems 'should be designed with the intention of augmenting, not replacing, human contributions' (Jarrahi, 2018) meaning that human and artificial intelligence can cooperate in symbiotic systems which utilize the comparative advantages of each. The IA approach alleviates fears that AI will replace the majority of workers, instead holding the positive stance that it will aid them in difficult and/or menial tasks, thus boosting their productivity. IA sounds promising in theory, but will it work in practice? Evidence from a report from Allied Market Research, answers in the affirmative, projecting that the global augmented intelligence market will reach $121.5 billion (£91 billion) by 2030. (Stephen, 2022) This gives businesses a strong reason to be confident in the embrace of artificial intelligence in the future of work.

The problem of bias in Artificial Intelligence:

However, there is still reason to exercise caution when embracing artificial intelligence in the business realm; namely, the problem of encoded bias. One must remember that Artificial Intelligence is not an independent entity, rather its creation and development are contingent on human inputs. The issue here is that humans are biased, both explicitly and implicitly. The result of this is that Artificial Intelligence has assumed, and now appears to perpetuate, human prejudices such as sexism and racism. Even if the intelligence augmentation approach (outlined above) is adopted, the question of whether the conscious human influence incorporated into it is enough to offset this inherent issue remains open, and the subsequent discussion will answer this question with a resounding "no". Even businesses that use AI to supplement, rather than replace, human inputs must recognise and respond to this concern directly.

The problem of ingrained bias in AI technology has been recognised in the academic realm for decades, for instance in 1998, Alinson highlighted that ‘a gendered vision of the world is inscribed in the technology of AI, albeit in a subtle way which must be uncovered or ‘de-scribed’’ (Alinson, 1998). Yet in the practical realm, this ‘de-scription’ (to borrow Alinson’s term) remains unachieved. Arguably, the first potential source of bias for any AI technology is the individual developers who create its algorithms. This risk of bias at this early stage of AI creation seems to have been mitigated by the development of specialized algorithms which are ‘created from parsing through large datasets of online information while having truth labels bestowed on them by *crowdsourced* masses.’ (Howard et al., 2018) (My italics) At first, this may seem to be a significant step in the right direction – personal prejudices appear to be avoided. But this is not the case, these algorithms ‘learn’ the implicit biases of society at large, and in doing so, ‘emphasize and reinforce these biases as global truth.’ (Howard et al., 2018) In this way, personal prejudice is not avoided, it is simply aggregated.

It is also worth noting that the majority of virtual assistants that are routinely utilised in various workplaces have one thing in common – ‘Alexa’, ‘Siri’, ‘Cortana’ – they all have the same default gender; female. This does not seem to be a mere coincidence nor is it of minor significance. Reflection on Schiller and McMahon’s 2019 article *‘Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence’*, leads to the realisation that routinely commanding[1] these gendered systems to complete menial tasks such as managing personal schedules, keeping track of to-do lists, and setting reminders, subliminally re-enforces outdated notions of gender-roles.

It is hard to argue that encoded bias is not a problem, after all, bias has countless, often detrimental impacts on human-flourishing (Howard et al., 2018). But businesses may ask whether this issue concerns them specifically (especially beyond the realm of corporate social responsibility). This question will be addressed in the following section through an examination of the potential impacts of encoded biases on businesses. This examination will also serve the purpose of further explicating the thinking behind the (above) assertion that even businesses with an intelligence augmentation approach to AI must address the concern of encoded bias directly.

The potential impact of biased Artificial Intelligence on businesses:

Reflecting AI’s varied and wide-ranging uses, the impacts of AI bias in the business realm are notably diverse. For instance, the effects of virtual assistants’ subliminal reinforcement of gender-roles will differ in scope, potency, and type from the impacts of the integration of driverless vehicles in the supply chain. This renders the task of fully unpacking these potential impacts a complex and comprehensive one. Therefore, due to the scope of this piece, attention will be focused on a case study: Unilever’s use of AI in its hiring process. This will enable certain impacts to be examined in sufficient depth.

[1] Note the significance of this verb

It is now common place for companies to use machine learning models trained on massive datasets of images to reduce costs in the hiring process. For example, the global leader in consumer goods, Unilever, has adopted the AI technology HireVue into their recruitment process. HireVue uses the above type of model to evaluate job applicants' video interviews; analysing facial expressions, body-language, and word choice. (HireVue, 2022) In light of the prior discussion, concerns could be raised that using this kind of technology in the hiring process risks racial and gender discrimination. However, HireVue reports that 'in just one year, the Unilever team saved over £1 million, reduced recruiting time by 75%, and *hired their most ethnically and gender diverse class to date.*' (HireVue, 2022) (emphasis added). This evidence, taken alone, points to the conclusion that the utilisation of AI in the hiring process can have a significant positive impact on a business's efficiency and profit, and that the biases that the academic realm highlights, do not seem to be as pervasive in practice as one may worry.

However, one should not jump to the conclusion that this negates the assertion that encoded bias constitutes a reason to exercise caution in the embrace of artificial intelligence in the business realm. This is because there remains counterevidence to HireVue's implication that the adoption of AI in the hiring process does not reduce diversity. This counterevidence can be divided into two strands. The first strand of evidence, demonstrated by Steed and Caliskan's finding that 'image representations learned with unsupervised pre-training contain human-like biases' (Steed et al., 2021), directly challenges the idea that the AI used in the hiring process does not have the potential to discriminate against applicants based on race and/or gender. Perhaps Hirevue helped to select an ethnically and gender diverse class, despite potential bias, simply because the pool of applicants was more ethnically and gender diverse than was historically the case.[2] The second strand of evidence, challenges the implication by suggesting that diversity could still be reduced in another way - pointing to other potentially encoded biases. For example, the type of learning algorithm used has been shown to associate thin people with pleasantness and overweight people with unpleasantness (Steed et al., 2021). Furthermore, the very idea that applicants should be assessed on 'word choice' could be accused of classism.

Now, it could be argued that AI is no more guilty of bias in the hiring process than a human would be, and therefore is the dominant option due to its superior (cost and time) efficiency. However, this argument can be easily overturned on various grounds. As a first response, it could be suggested that the rise of initiatives such as "unconscious bias training" help humans to recognise where such biases are influencing their decisions. The current AI systems used in the hiring process do not exercise this type of second-order evaluation. On top of this, in analysing factors such as body language and word choice, AI may miss some of the nuances of human conversation. So, while the AI technology may not be directly discriminating based on social factors, such as class, it may do so indirectly. For example, passing off a candidate with a less sophisticated dialect as being less knowledgeable. Arguably, a well-trained human employer would be able to differentiate and navigate such nuances.

This points to the conclusion that the embrace of AI in the hiring process, and perhaps in the business realm as a whole, does have a negative impact on equality and diversity in the workplace. (As explained above, the evidence against this conclusion seems to be

[2] Further research is required to assess whether this initial idea is reflective of the true situation in the case study.

outweighed). If this is the case, its use will hamper businesses' future success, as workplace diversity has been shown to generate innovation, enable professional growth, and foster better decision-making. (Stahl, 2021). This makes it clear that businesses must tackle the issue of encoded bias in AI directly, especially if they wish to use AI in processes such as hiring.

How should businesses respond to the problem of bias in Artificial Intelligence?

To recapitulate, it has been established that the problem of encoded bias has the potential to negatively impact businesses, specifically by reducing equality and diversity in the workplace. In order to mitigate these potential negative impacts, businesses must tackle the issue of encoded bias directly, but the question is – how? This piece will suggest that businesses can, and should, persist with the embrace of AI in a more conscientious way, utilising AI audit procedures in order to ensure that the technology they adopt is fit for purpose.

Having identified and quantified AI ethical failure, researchers from the University of Oxford's Saïd Business School (Floridi et al., 2022) developed an ethics-based audit procedure intended to help avoid failures such as intrusion of privacy, lack of explainability of decisions made, and (key to this discussion) algorithmic bias. This tool, called capAI (conformity assessment procedure for AI systems), adopts a process view of AI systems; evaluating current practices across the five stages of the AI life cycle: design, development, evaluation, operation, and retirement, ascertaining whether the system in question is legally compliant, technically robust, and ethically sound.

Not only will such audit processes help businesses to identify where AI systems are unacceptably biased, and therefore avoid the negative consequences of this. It appears that they will soon be widely legally required, so not only *should*, but *must* be adopted. For instance, the draft EU Artificial Intelligence Act (AIA) explicitly sets out a conformity assessment mandate for AI systems, and the 2022 amendment to the 2019 US Algorithmic Accountability Act, also mandates AI impact assessment. (Floridi et al., 2022) Therefore, one is fully persuaded that in order to best utilise AI in the future of work, businesses need to make use of audit procedures such as capAI.

Conclusion:

In conclusion, there is a strong reason to exercise caution when embracing artificial intelligence in the business realm. Previously, it may have been thought that this reason was the fear of machine dominance, however the trend towards an intelligence augmentation approach to AI (which aims to integrate human and artificial intelligence into cooperative symbiotic systems) seems to have answered this concern. Today, the reason why businesses must exercise caution in their adoption of AI is the problem of encoded biases which run the risk of reducing equality and diversity in the business realm. This is especially true if unacceptably biased algorithms are used in areas such as hiring and recruitment – a usage trend which is already evident. It has been demonstrated that this issue should be tackled directly, specifically through the utilisation of AI audit procedures. The strength of this conclusion is bolstered by its apparently increasing political support, as reflected by recent changes in EU and US legislation. (Floridi et al., 2022).

References:

Alinson, A., (1998) *Artificial knowing: gender and the thinking machine* Routledge

Floridi, L., Holweg, M., Taddeo, M., Silva, J., Mökander, J., Wen, Y., (2022) *capAI - A Procedure for Conducting Conformity Assessment of AI Systems in Line with the EU Artificial Intelligence Act 2022*

HireVue (2022) *Unilever finds top talent faster with HireVue assessments* HireVue available at https://webapi.hirevue.com/wp-content/uploads/2020/09/Unilever-Success-Story-PDF.pdf?_ga=2.69252048.1719659662.1668379873-1105663879.1668379873

Holweg, M., Younger, R., Wen, Y., (2022) *The Reputational Risks of AI* California Management Review, Volume 65, Issue 1, Summer 2022

Howard, A., Brostein, J., (2018) *The Ugly Truth About Ourselves and Our Robot Creations: The Problem of Bias and Social Inequity* Science and Engineering Ethics volume 24, pages1521–1536

Jarrahi, M., (2018) *Artificial intelligence and the future of work: Human-AI symbiosis in organizational decision making* Business Horizons, Volume 61, Issue 4, July–August 2018, Pages 577-586

Mirbabaie, M., Brünker, F., Möllmann Frick, N.R.J. et al. (2022) *The rise of artificial intelligence – understanding the AI identity threat at the workplace.* Electron Markets 32, 73–99. available at https://ezproxy-prd.bodleian.ox.ac.uk:2102/10.1007/s12525-021-00496-x

Rai, A, Constantinides, P & Sarker, S (2019) *'Next Generation Digital Platforms: Toward Human-AI Hybrids'*, MIS Quarterly, vol. 43, no. 1, pp. iii-ix.

Schiller, A., John McMahon, J., (2019) *Alexa, Alert Me When the Revolution Comes: Gender, Affect, and Labor in the Age of Home-Based Artificial Intelligence,* New Political Science, 41:2, 173-191

Stahl, A., (2021) *3 Benefits Of Diversity In The Workplace* Forbes available at https://www.forbes.com/sites/ashleystahl/2021/12/17/3-benefits-of-diversity-in-the-workplace/?sh=ce49aae22ed2

Steed, R., Caliskan, A., (2021) *Image Representations Learned With Unsupervised Pre-Training Contain Human-like Biases* arXiv:2010.15052v3 [cs.CY] 27 Jan 2021

Stephen, A., (2022) *The future of augmented intelligence* Business: the next 25 years available at https://www.sbs.ox.ac.uk/oxford-answers/future-augmented-intelligence last accessed: 10/11/2022

West, S.M., Whittaker, M. and Crawford, K. (2019). *Discriminating Systems: Gender, Race and Power in AI*. AI Now Institute. Retrieved from https://ainowinstitute.org/ discriminatingsystems.html.

Zhou, L., Souren, P., Demirkan, H., Yuan, L., Spohrer, J., (2021) *Intelligence Augmentation: Towards Building Human-Machine Symbiotic Relationship* AIS Transactions on Human-Computer Interactions; Atlanta Vol. 13, Iss. 2, (Jun 2021): 243-264.