Martin Christoph Preis
(22214417)
21.03.2024

Risk Assessment or Risky Business?

How an ethical dilemma challenges the social Impact of AI

Clusifine.ai, a transformative social finance startup, has positioned itself at the forefront of revolutionizing the microcredit application and permission processes through the development of cutting-edge algorithms.

The success story of Clusifine.ai

It's founder, the Washington native Sophia Patel previously made a fortune as a co-founder of a fintech startup that developed a solution for venture capital firms to calculate the risks of startup investments. After a personal crisis, she decided to invest her money and energy for a good cause and founded the impact startup Clusifine.ai.

With the mission to make microcredits more accessible and cost-effective, Clusifine.ai developed an algorithm that works in its fundamentals similar as the solution of Sophia's previous venture but is based on variables that are more accessible for the new target groups of microentrepreneurs in underprivileged regions of developing countries. On the front end, grant seekers like microentrepreneurs or collectives can input relevant data including demographic information, business details, financial statements, and socio-economic indicators into an intuitive interface. In the back end, is a complex algorithm, which is trained with historic data from banks and financial institutions from the respective countries or regions. Drawing from both the inserted and the historic data, the algorithm calculates a total risk score between 0 (very high risk) and 100 (no risk) and divides the scoring in 5 categories:

1. Credit History
2. Financial Health
3. Business Performance
4. Market and Industry Factors
5. Demographic and Socio-Economic Factors

This risk assessment helps potential credit providers to decide whether or not to approve a credit. An assessment that would currently take several hours of work can now be completed in just a few clicks and much more accurately than before.

Clusifine.ai embarked on a rigorous testing phase to evaluate the efficacy of its algorithm. The testing took place in Peru, where the results demonstrated significant potential for simplifying and streamlining the application and permission processes, which have historically been plagued by high costs and time-consuming procedures. The development and testing were accompanied by researchers in the fields of development studies, economics, and data science. In the scientific papers published

after the test phase, all the authors, without exception, were very positive about the results and the potential of the Clusifine.ai solution: On the one hand, the application could significantly reduce the time and costs associated with each microcredit, leading to an increase in the number of microcredits with lower interest rates for the beneficiaries. On the other hand, and this is the more important message, the software could reduce the risk of lending and make it much more calculable. This in turn would make it much more attractive for potential financiers and investors to channel capital into the microfinance sector. With this fresh capital, the volume of microcredit lending itself could grow significantly, and more loans could be made available to help people in underprivileged regions out of poverty and stimulate economic growth.

Andrea Osei and the second testing phase

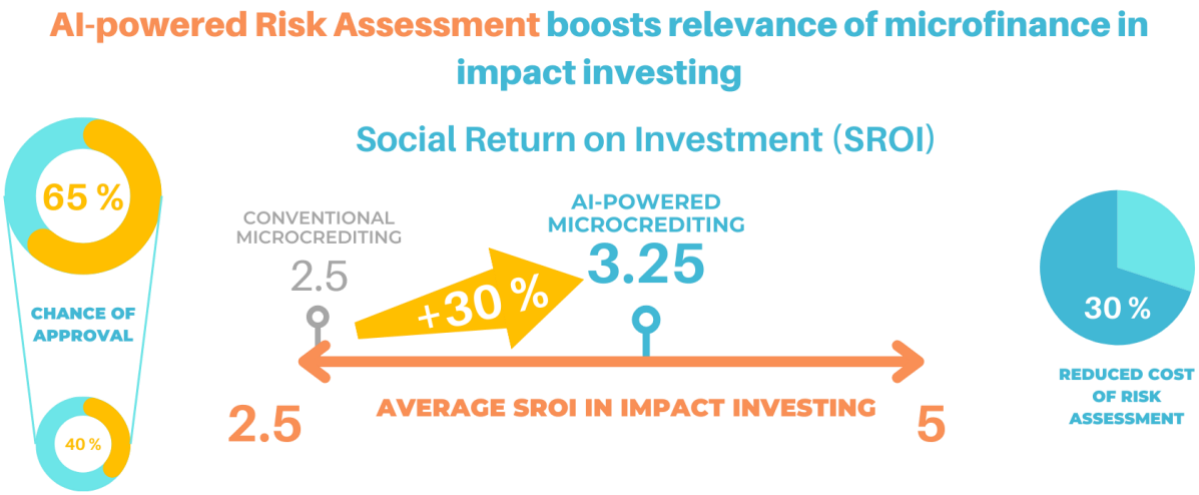
To spearhead the implementation of the algorithm in Nigeria and Sri Lanka, Clusifine.ai brought onboard a young and talented data scientist named Andrea Osei. As soon as she was onboarded, she took over the lead of the test phase in Sri Lanka. The tests in the other two countries were already well underway and with the support of her colleagues there, she was able to familiarize herself quickly. The procedure developed by the testing team was basically correct, but Andrea found the tests to be very rudimentary. She had learned a few test methods during her studies and used her knowledge to perform additional test methods, including a geographical mapping of the data points and a clustering by demographic characteristics, in addition to the intended tests. During the very first test run, she noticed several anomalies - e.g. a visible geographical distribution of lower scores in the north of the country. The results irritated her and, after reviewing the methodology several times, she sought an informal exchange with her colleagues to find an explanation. However, the reaction of her colleagues ranged from irritation to disinterest. They pointed to possible errors in her methodology, as she had deviated from established procedures, and insisted that nothing of the sort had been observed in her tests. Andrea, however, could not let up, as she was convinced of the correctness of her approach. She searched the demographic clusters for a data point that could be the cause of the anomalies and it didn't take her long to find it.

Andrea Osei’s journey to impact and success

Born in Ghana and having relocated to the United Kingdom at the age of 12, Andrea's upbringing was influenced by her parents' work as medical technology researchers at Ghana's prestigious Korle Bu Teaching Hospital. Her parents' subsequent move to Glasgow, Scotland, where her father accepted a professorship at the esteemed Strathclyde Institute, laid the foundation for Andrea's academic journey.

Andrea was always interested in technology and therefore pursued a degree in Computing Science and Software Engineering at the renowned University of Glasgow. During her Masters, she interned in the fintech department of a London-based consultancy. They offered her a high-paying entry-level position upon graduation, and she accepted. She was very pleased with her first month at work, her work was interesting, and she developed a quick comprehension for financial topics. She could also afford a great standard of living, so London really had a lot to offer her. But after

less than a year, Andrea noticed that she was missing something, and she also knew that her parents weren't really excited about what she was doing because they had a generally negative attitude towards the financial industry.
During that time, she got captivated by the philosophies of Andrew MacAskills' Effective Altruism movement. After accidentally stumbling over his book “Doing Good Better”[1] in a bookstore in her hometown, she was convinced that with her computing skills she could achieve a substantial social impact. Although her parents had also dedicated themselves to the common good through their work in medicine, she was never very strongly influenced by them in her career choice. She was not indifferent to ethical issues, but she regarded volunteering and social work more as symbolic activities to salve one's own conscience, while simply ignoring these issues outside of these activities. She also missed the transformative aspect that she now found in Effective Altruism. The approach of not simply getting a good feeling by helping, but generating the greatest possible impact based on scientific findings, in the best case in one's own career choice or through entrepreneurship just made a lot more sense to her. Driven by this philosophy, Andrea began to engage with companies and organizations that put social impact at the heart of their activities and develop solutions to major social problems. This is how she became aware of Clusifine.ai through a post by the Falling Walls Foundation, in which Sophia presented their estimated social Impact:



The post showed how Clusifine.ai has made significant progress in the biggest problem of microcredit, namely risk assessment. This not only led to improved profitability, but the simplified approvals also significantly increased the social impact, as measured by the social return on investment (SROI). Microcredits would become a much more attractive option for impact investing thanks to increased impact combined with lower risk. Put simply, this means that more people are given the opportunity to lead a more self-determined life.

[1] MacAskill, W. (2015). *Doing good better: How Effective Altruism Can Help You Make a Difference*. Penguin.

As a Ghanaian in the diaspora, she was convinced by the start-up's goals and felt that she was a good fit for the company. Andrea saw Clusifine.ai as an ideal platform to leverage her expertise and create meaningful impact by contributing to the upliftment of underprivileged regions.

In Clusifine.ai, she saw the potential to provide underprivileged individuals with access to capital as a catalyst for breaking the cycle of poverty, which led her to join the company. From the first day on, Andrea liked the young, dynamic and impact driven nature of the company. Even though the social impact was always in the foreground, everyone was as professional and success oriented as she knew and appreciated from her previous work experience. She was convinced that with a job at Clusifine.ai she found the opportunity to effectively create social impact and sent out an open application the same day. Apparently, she had chosen a good time, because the answer was not long in coming. She received an offer for a position in Business Development and during her job interview, Sophia Patel even joined in for a moment to get to know Andrea. With her energy and determination, she immediately captivated Andrea and gave her the feeling that they both had the same goals in mind. And so, Andrea didn't hesitate for long before accepting the position.

Sophia Patel’s hype in the social impact bubble

Andrea first became aware of Clusifine.ai because the company received a huge amount of media attention out of nowhere. A few months before she was hired, and buoyed by the success of the testing phase, the Clusifine.ai team had embarked on a highly confident communications strategy to gain international visibility and attract investors. The core message they wanted to spread was that Clusifine.ai was going to "revolutionize microcredit". In a recent keynote speech by Sophia Patel at TEDxPaloAlto, she described the relevance of Clusifine.ai as follows:

One of the greatest challenges for the global poor is access to capital! In the 1970s, Mohammed Yunus invented microfinance to address this challenge and revolutionized the financial industry to become part of the fight against poverty. Microfinance was a game-changer! However, it did have one problem: the assessment of risk causes great costs and is an enormous obstacle to the success of the entire microfinance sector. But with AI, we now have the chance to overcome this obstacle! And 2024 could be the year in which Clusifine.ai revolutionizes the finance industry for a second time, enabling microfinance for every poor person around the world!

The campaign had great success, Sophia Patel and Clusifine.ai went viral in the philanthropy and global development communities, with presence in magazines, invitation to the World Economic Forum and the Bridging Gap Foundation.
In their first funding round, Clusifine.ai presented the results of the testing phase to its prospective investors, including philanthropic funds and impact investors. The impressive outcomes not only solidified investor confidence but also led to increased funding. This newfound financial support enabled Clusifine.ai to expand its operations and directly apply its innovative tool in three different countries of three different world

regions: Peru in South America, Ghana in West Africa, and Sri Lanka in Southeast Asia. In a pre-selection process, countries were selected that a) represented one of the three regions of the global South (Africa, Asia, Latin America), b) had only a moderate level of corruption (> 30/100 on the perception corruption index)[2], and c) promised significant economic growth potential when looking back at the last 30 years[3]. From this shortlist, the three countries mentioned above were selected because of the potential for public-private partnerships to enter the market. By targeting these diverse regions, Clusifine.ai sought to accelerate its impact on a global scale, so that, if successful, it could create a “big bang” in the microfinance branch, like Sophia was hoping for.

Racial bias in AI

After examining the demographic clusters, Andrea was able to identify a cause for the anomalies in the assessment. While in the majority of the loan applications analyzed, Singalese was the native language of the applicants, about 15% of the applications were from native Tamil speaking individuals. In a comparison, Andrea was able to clearly show that, with comparable profiles, the Tamil applications had a significantly inferior risk assessment.

After stumbling over the racial bias in the Algorithm of Clusifine.ai, Andrea spent her weekend in London’s Anthropology Library. She was convinced to find the root of the problem within Sri Lankas ethnographic context. So she started digging and learned about the long-standing conflict between Sri Lanka's two largest population groups, the Tamils, and Sinhalese. The conflict dates back to the time of colonization by the British Empire and resulted in various violent conflicts and a bloody civil war from 1983 to 2009. To this day, there is strong tension between Tamil and Sinhalese Communities. Because the Sinhalese were able to make amends with the British authorities during the colonial period, they were able to secure government posts and continue to enjoy political and economic dominance to this day[4]. This is where Andrea found the explanation for the AI's bias: Because Clusifine's algorithm uses historical data to make decisions about the creditworthiness of individuals. And the historical and ongoing oppression of the Tamils is reflected in various ways in parts of that very data:

1. **Economic inequalities:** Many Sri Lankan Tamils lost their owned land and were excluded from employment in the public sector. This led to a structural economic inequality in favor of the Sinhalese population and Tamil business owners being statistically less likely to become economically successful.

---

[2] Ortiz-Ospina, E., Roser, M. & Our World in Data. (2024, 18. März). *Corruption*. Our World in Data. https://ourworldindata.org/corruption

[3] Our World in Data. (2024). *Gross domestic product (GDP)*. 1950 to 2019. https://ourworldindata.org/grapher/national-gdp-penn-world-table?tab=chart&country=LKA~GHA~PER

[4] Stokke, K. & Ryntveit, A. K. (2000). The Struggle for Tamil Eelam in Sri Lanka. *Growth And Change*, *31*(2), 285–304. https://doi.org/10.1111/0017-4815.00129

2. **Financial exclusion:** Because of discriminatory practices, Tamil businesses and individuals have been less likely to have been granted credits.
3. **Educational exclusion:** Sri Lankan Universities admission processes favored Sinhalese applicants, leading to lower average education levels of the Tamils.
4. **Reduced public investment:** The Sri Lankan government reduced investment and public spending to Tamil regions leading to disproportional economic stimulus in Sinhalese regions.
5. **Linguistic dominance:** As Sinhalese was defined as official language in Sri Lanka, data from Tamil speaking businesses is underrepresented.
6. **Cultural norms and practices:** Historic data represent the perspectives, experiences, and preferences of the dominant Sinhalese culture and inherits stereotypes and cultural assumptions in disfavor of the Tamil.

For Andrea it was obvious that there were multiple ways and possibilities how the historical data could lead to a racial bias in Clusifine.ai's predictive model. The irrefutable conclusion she drew was that the problem she discovered was not a minor bug, but a profound underlying problem in the way Clusifine.ai's algorithm is trained.

Breaking the news

Alarmed by her discovery, Andrea sent a message to her supervisor, Jean Satigny, early on the next morning outlining the problem. *"During a viability check of the test results, I found a clear bias against people from a certain demographic group. I see clear indications that this bias is not due to a bug, but is reflected in the cross section of the training data used,"* she wrote.

It took almost two hours before she received a reply to her message, which for Jean was highly unusual. Jean wrote: *"If your suspicions are correct, this is a very serious matter. I have informed the Leadership Board and we have scheduled an emergency meeting for tomorrow at 09:00. Until then, please try to gather your data in as much detail as possible*". Andrea, already worried, was at that point completely shaken. She started working late into the night and even forgot about the dinner plans she had with a close friend that night. She went to bed but couldn't sleep because she couldn't stop thinking about meeting the next day. How will Sophia react? Is it possible that she already knows about the problem? How will the other Leadership Board members react? She has been to two leadership meetings and found Sophia to be quite predominant. There was always very little discussion within the board, much less than in her previous work. She began to fear that the discovery might have an impact on her own job and mentally reviewed her beginnings at Clusifine.ai.

The change from a profit-oriented to an impact-oriented company was a touching experience for Andrea. After several months of feeling empty and disengaged, she gave herself the courage to do something about it and find a meaningful career, following her newfound ideas of effective altruism. She remembered how excited she was when she told her parents about her new profession, and they rejoiced with her, proud that their daughter would work not only for the common good, but even for her country of origin. For the first time, she wondered if her work was only having a positive

effect, or if she was also responsible for side effects of Clusifine.ai that she had not thought of. These thoughts did not leave her for a long time, until at some point she fell into a troubled sleep.

At the leadership board

The next day she wakes up really unrested. From the beginning of the meeting on she is a little bit scared and intimidated from the attention that is resting on her. She was never a person who wanted to be in the center of attention and is really discomforted from the whole situation. But she took a heart and decided to stick to the presentation she prepared and get it over with. Jean opened the meeting by describing how he received Andrea´s message and that he immediately took the issue very seriously and therefore called this emergency meeting. He then kindly asked Andrea to present what she had found out. She started her presentation by describing in detail the evaluation she was working on, the data that was involved, the methodology she used and the tests she conducted.

Andrea described how she became aware of a strange test result that prompted her to investigate other testing scores that showed a similar anomaly. After examining around 50 scores, she was able to identify a pattern of people from certain regions in the north of the country achieving structurally lower results despite having similar profiles. In the second part of her presentation, she looked at the history of Sri Lanka. She summarized her findings on the socio-political situation of the country and showed how these are reflected in the training data used by Clusifine.ai. Andrea used the last part of her presentation to illustrate how deeply these discriminatory tendencies would feed into the algorithm and that this would significantly affect the test results. With that, she ended her presentation and deliberately left out any recommendations for action, as she did not want to patronize the management on such a central issue.

After the presentation ended, it took a moment for Sophia to take the floor. She first thanked Andrea for the insightful information and complimented her on her vigilance during the evaluation. She even went further and described how well she had familiarized herself with her tasks in such a short time and that she would like to see her grow with the company. Andrea was slightly touched and delighted because she knew that she was indirectly being offered a promotion. Sophia gave her the feeling that at least she didn't have to worry about her job and that reassured her a little.

But after that the Clusifine.ai CEO diverted the conversation and began to speak about the company's current situation. She shared insights about the investor discussions she is currently in and about to finalize, her conversation with the Chairman of the U.S. Agency for International Development who is interested in Clusifine, and meetings with key international philanthropists. She then showed a graphic with the milestones that the start-up wants to achieve in the coming months:

Sophia clarifies that it would not be an option to fundamentally intervene in the training of the algorithm at this point in time. “All the relevant stakeholders are expecting the first launch of Clasifine.ai in Peru and would be very concerned if we were to go back to testing now. This could bring the whole company in risk when we would get the image of being unreliable. In addition, the preliminary talks for the phase A funding phase were already in progress. This new funding could further enhance our growth, especially for the launch in Ghana, where we are already making plans with our local partners. All of this could suddenly be up for grabs again.

Jean speaks reacts and notes that it would be "extremely difficult, if not impossible" to eliminate the bias after the launch. “What motivates our partners to work with us is the impact we are promising. If we let any doubt about that come up, this could cause us far-reaching problems - we have to take this seriously!”
To move the discussion in a more constructive direction, Andrea suggests: "We could form a task force dedicated to the problem. Together with a team and some time, we may be able to find an adequate solution."

Sophia reacts with tension: "We are on the verge of a business breakthrough. We can't afford to step back at this point. Our current plan is really ambitious, so we have to show internal and external unity. We can't get any more of our staff involved in this.” She pauses for a moment, then turns to Andrea: "Andrea, as you know, preparations with our partners for the launch in Peru are about to begin. Please take the next two weeks to find a solution that will at least help us narrow things down – at least until the launch. In two weeks, we will meet in the board again and discuss our next steps. And please, keep all of this confidential.”

Andrea’s strategy
Andreas' temporary feeling of reassurance had now completely dissipated. After the meeting she canceled all her appointments for the next two weeks, then took a step

back and thought about how to approach the situation. She realized that what Sophia was asking her to do could not be done correctly. The test results could be corrected manually in the most obvious cases, but a real solution to the problem would take much more time. This solution would probably require intensive preprocessing of the training data, which would be impossible within the current cost model and Clusifine's scaling plans. She was relieved that Jean agreed with her diagnosis. She realized - Clusifine had a deep-seated problem, and everyone involved was well aware that Sophia was not willing to react appropriately.

In order to be able to discuss her thoughts with someone, Andrea asked Jean for a meeting, which took place a few days later. She had approached Jean because he had both initiated the emergency meeting and was the only one in the meeting to represent her position and point out the importance of the problem. That day, however, he no longer seemed as convinced as before. She shared her thoughts with him and asked him what he would advise her to do. He replied that he still thought the position he had taken in the meeting was correct. Racial bias is a serious problem and a social startup with flat hierarchies like Clusifine.ai should provide the internal governance-structure to deal with such a problem more transparently. However, he would also understand Sophia, who views the problem from an entrepreneurial perspective. Clusifine.ai would be very vulnerable before they have established themselves, if something were to leak out now, it could bring the whole company down. From an economic and social point of view, that would be the worst possible outcome. A scandal could not only bring down their company but damage the entire microfinance industry and ensure that AI solutions in this area are permanently discredited. Jean then concluded by saying that he could not tell her what to do, but that she should consider both. He then left the call by claiming that he had to jump into a next meeting.

It was now clear to Sophia that she was on her own with this problem and needed to develop a plan as quickly as possible. She considered her options: She could, just as Andrea had asked her to, propose a surface-level adjustment to the algorithm that would appear to address the racial bias. This would de-escalate the situation and buy more time for the management until after the launch, when Clusifine.ai would be less vulnerable and have more financial capacity. Another option could be to develop a strategy that prioritizes transparency and accountability in a genuine effort to address the root causes of bias and convince Sophia that this could open up a broader solution space to develop innovative solutions. The conversation with Jean made her realize once again what she subconsciously already knew. She is at a crossroads and the strategy she has to present to the board will have a huge impact on Clusifine.ai, her own career and potentially the entire microfinance industry that reaches millions of people in the Global South. So, what should she do?

Tasks

a. Put yourself in Andrea's perspective: Discuss in which type of normative tensions she finds herself in? Which ethical considerations could she make?
b. What type of paradoxical tensions are present within Clusifine.ai? Please elaborate.
c. After you have analyzed the problem, please put yourself in Andrea's perspective again. Assign the two decision paths that Andrea mentally describes at the end of the case as an instrumental or a integrative strategy and explain your decision. Decide on one path and develop a strategy for how she could deal with the situation.