

Good news in the field of Data Science

Laura Solà Garcia

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

laura.sola.garcia@estudiantat.upc.edu

Joan Ot Vidal Camps

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

joan.ot.vidal@estudiantat.upc.edu

Blanca Unanue Gamba

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

blanca.unanue@estudiantat.upc.edu

Ramon Ventura Navarro

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

ramon.ventura@estudiantat.upc.edu

Aida Torres Company

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

aida.torres@estudiantat.upc.edu

Maria Tubella Salinas

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

maria.tubella@estudiantat.upc.edu

Marcel Tomàs Bernal

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

marcel.tomas@estudiantat.upc.edu

Laia Villagrasa Umbert

Grau en Ciència i Enginyeria de Dades

Barcelona, Spain

laia.villagrasa@gmail.com

Enric Rosero Pardo

Grau de Filosofia
Universitat de Barcelona

Barcelona, Spain

enric.rosero@gmail.com

Elsa Verdera Reig

Grau de Filosofia
Universitat de Barcelona

Barcelona, Spain

elsavereda@gmail.com

Abstract-

This paper aims to explore the views of data scientists on the positivity or negativity of their work, expose the more positive aspects of the application of data science in various disciplines, and define what paths should be taken to maximize the positive impacts of technology in the future. Firstly by doing research on the perception that data scientists have over whether the impact of technology is generally good or bad by compiling their views on the impact of AI in the SDG (Sustainable Development Goals) and analyzing their negativity bias. Secondly by exploring a case study on the application of AI-based technology with the intent of making wars less brutal and protecting civilians. Thirdly by analyzing through possible future news and inventions both what steps are considered essential to take by future data scientists to ensure a world filled with good news regarding their work and what disciplines require the most attention put into them always with the objective of improving the world as a whole. Finally, six prevalent topics previously explored in class are exposed accompanied by possible solutions thought of by those who worked on them in an attempt to provide some objective to follow to gain a future full of good news.

Keywords- Good, impact, pessimism, optimism, AI, future, news

I. INTRODUCTION

It is in human nature to focus on bad news [1][2] no matter the topic, but this is especially relevant in the world of AI and data science, as technologies are mistrusted, information is scarce or not very accessible and incidents such as data leaks are not uncommon.

It is in this environment in which data scientists have to develop their projects assumedly with good intentions. This paper, therefore, aims to explore what good there can be in the application of AI, how that goodness is perceived by those who will work on it, and how to ensure a world where the amount of such good news is maximized.

II. THE MEANING OF GOOD

In order to assess and improve an attribute, no matter which, it is indispensable to be based on a solid definition. When talking about the goodness of AI, this is difficult, as such a concept can be explored through many lenses. Therefore firstly we aim to define it to have a more solid base from which to expand this paper and its contents.

Philosophy, often considered the foundational discipline of sciences and human thought is the way through which to explore it. Firstly, as many philosophers such as Heidegger or A. Jonas have

argued, technology is not neutral itself, but rather affected by the values present in society and within its creators, therefore it is impervious to analyze the context they are developed in to judge their goodness.

Secondly, it can be judged by the procedures used, how, where, and in which way they are developed. How are the materials to create it extracted? Is the public privy to and a participant in the process? Any inequalities in this development are reflected in the ethics of the technology. Thirdly, consequences play a great part in assessing the goodness of an AI application, but not only those often taken into account in news, good short or mid-term consequences, but also those encompassing all time, people, and resources be they grievances or favors. This also includes responsibilities with their many degrees and tones, what the relationship will be between the agents and their unequal capabilities and knowledge, and what that might cause.

In summary, analyzing the goodness of the application of AI has to be a careful and pragmatic yet not unnecessarily fearful process full of questions centered around “What life does this technology generate and permit?”. If technology is humanity’s answer to maintain, preserve, and improve life, then its inception is founded on keeping and ameliorating the quality of life, an improvement of eudaimonia, the possibilities of being and existing in this world for all and each human and non-human creature. All good technology should therefore strive to give access to a good life, especially to those less fortunate.

In conclusion, ethics should be used to analyze AI, from its inception to its application and methods, questioning how it should be used and developed taking into account the biological and social impacts it might have. Keeping up this constant dialogue so that they bring society to true progress benefiting everyone and not only the privileged few, always cautious of potential misuses.

III. PESSIMISM AND OPTIMISM

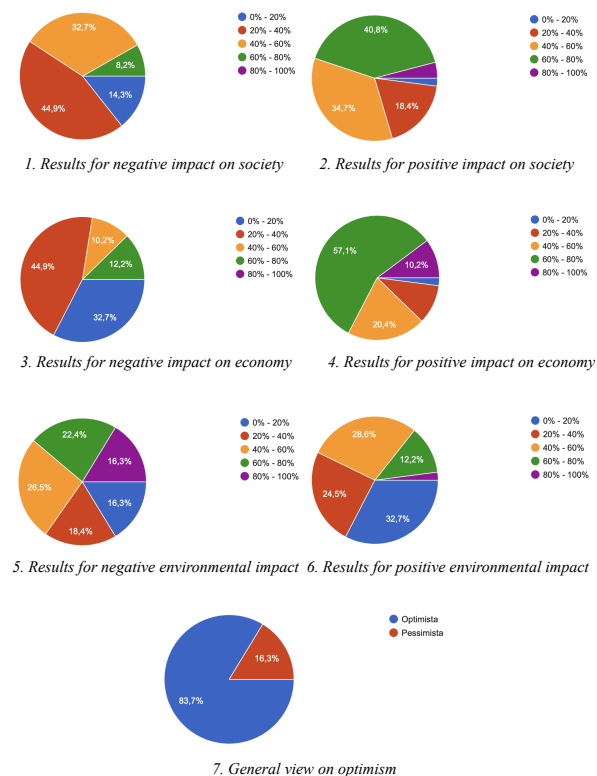
Existing in the data science space and researching the many ethical dilemmas that have been and could be caused by the developments in this field, one could come to the conclusion that it might do more harm than good or generally fall into pessimism around the subject. This is why a small study was conducted to get an idea of how prospective data scientists feel about whether their future work will better the world

or not.

More specifically the topic analyzed was their views on the impact of AI in the SDGs [3]. The Sustainable Development Goals are a set of objectives, split themselves into different feats, which point to the environmental, social, and economic developments that should take place in order for the evolution and existence of human society to be sustainable. Due to the undeniably huge impact AI has had in the world of technology and therefore in human life, it is imperative to ponder over whether it will set us closer or further away from the SDGs.

According to a study conducted in 2020 [4], the potential percentage of objectives for which AI will be an enabler in the environmental, social, and economic subgroups are, respectively, 93%, 82%, and 70%. In comparison, the percentage of objectives that will be inhibited is 30% in the environmental subgroup, 38% for the social one, and 33% for the economic goals.

The following charts illustrate the results obtained from a poll conducted on a group of 49 data science students..



What was demonstrated through this activity is, on one hand, that most data science students’ assessment of the negative was generally realistic. Specifically,

more than half of the polled correctly placed the negative percentage of social effect to be between 20% and 40% with the majority of the rest assuming it to be between 40% and 60%. The majority also concluded the percentage of negatively affected economic goals to be between 20% and 40% with the remainder of the polled mostly guessing it to be more positive, between 0% and 20%. The biggest difference was observed in the environmental percentage which most people surmised to be between 40% and 60% when in reality it is between 20% and 40%.

On the other hand, their assessments of the positive impacts were mostly lower than the actual value. Particularly, regarding the social aspect, the majority decided it to be between 40% and 80%, and more surprisingly almost half the students placed the positively impacted environmental goals to be between 0% and 40%, with nearly no one correctly assessing it to be between 80% and 100%. Most of the polled considered the positive impact on economic objectives to be between 60% and 80% which matches with the results of the study[4].

Conclusively, the general belief from the students was that their view was an optimistic one. In contrast, from the results of the poll it could be stated that they veered towards pessimism, overestimating negative impacts and underestimating, in some cases vastly, the amount of positively impacted goals. This was especially apparent in regards to the environmental objectives, which could be due to the fact that the impact of AI on pollution is a commonly discussed issue but the way it can help revert desertification, preserve the health of ecosystems, and improve energy efficiency is not often talked about.

After this small study, another test was given to the students, this time to make them aware of the negativity bias. This bias is a cognitive one that causes adverse events to have a more significant impact on our psychological state than positive events. This effect has been proven through experiments such as the one led by Claire E. Robertson [5]. It is necessary to point out, though, that negativity bias doesn't exist in a void, there also exists a positivity bias, a tendency to form optimistic theories about reality and in consequence a positive-negative asymmetry, although, if put in a scale we still tend to give more importance to the negative side.

To conduct the test, the class was shown a few

headlines 9 of which were good news and 8 of which were bad. Afterward, they were told to write down the three ideas which stood out the most to them and rate the impact they thought the news would have on a scale from 1 to 10.

The results were that 45 of the headlines remembered were positive and 73 negative, in fact, out of the top four most voted, 3 had a negative implication and only one a positive one. Furthermore, the impact of the negative news was judged to be, in mean, higher than the positive ones. Thus demonstrating further that the class reacted more pointedly to "bad news" and was not impervious to negative bias and pessimism.

IV. CASE STUDY

Seeing how pessimism appears to be a tendency, a debate was proposed which aimed to discuss whether new technologies could be used for good or reduction of negative impacts even in desperate situations.

One such undeniably negative event is war, considering the amount of lost lives, destroyed societal structures, and material damage it causes. Therefore this is the topic the discussions were centered on, to aid it, two current applications in different directions were introduced.

A. AI FOR TARGETING

Firstly the use of AI for the targeting of military objectives was presented as a debate topic through two news articles. One [6] which explained how an AI system named LavanderAI was utilized to identify 37,000 targets, making the process easier and more precise to the given information but, at the same time dehumanizing those killed. The other [7] which exposed how an AI platform by the name of the Gospel has been recently deployed. The Gospel permits the identification of about 100 targets a day, whereas the usual rate used to be 50 a year and its main objective is "inflicting great damage to the enemy and minimal harm to non-combatants" [7]. It is also explained, though, that many external sources doubt the efficacy of the platform and point out how it could be used as a tool to run a "mass assassination factory" or make humans develop "automation bias"[7].

The students were prompted to argue whether they agreed with this use of technology both from the point of view of the military and of civilians. Taking into

account that the targeting system had a 90% accuracy rate and that the one who ordered the kill was a human, aware of the potential collateral damage of the bomb and able to throw it anywhere. The conclusions they reached were mostly in favor of the use of AI, claiming it would reduce the time, resources, and number of deaths wars take, that killing specific targets is better than indiscriminate death, and that it might both take responsibility away from the military and make civilians feel safer. In addition, they argued that this is a dog-eat-dog world and therefore if they didn't use this technology someone else would and that they would be on board if the killing was regulated by a committee. The few who were against it were mostly civilians, who found their rights weren't being protected and that potential regulations would surely be ignored in times of war.

As can be seen, most of the arguments took into account ethical criteria such as vulnerability, as they attempted to protect the least fortunate, such as defenseless civilians, or reversibility and universality, as they considered what would happen if they were to receive the action or everyone had access to this system, reaching the conclusion that not having it would put the country into great disadvantage and if the whole world had it, wars would become bloodier. They also pondered the transcendentalism of it, how killing and objective would also doom their entire family, but that was often overshadowed by utility as they ultimately considered the maximum number of people benefiting from the action to be most important so even if 10% of the objectives were innocent they assured many benefited from the 90% that weren't being murdered.

Next, they were asked if they would still agree given the bomb could be thrown into houses or if they would only target military bases, with the awareness that the first option would cause more deaths of innocent humans but cost less in resources than the second. Students were more torn in this case, with those in favor of bringing the war to the cities in hopes of shortening the war, arguing that killing the military leaders would lead to an earlier end for the war and as a consequence the death of fewer innocent lives. Those against it gave arguments such as how this would move the war front to cities and how the death of innocent lives would lead to fast escalations of the war efforts.

In this case, the objective of the setting was to make

them ponder between their utilitarianism and the sense of vulnerability of the victims. As a result, most arguments were consequentialist in nature, valuing a victory in war and the potential saving of many lives more than the assured death of a few innocent civilians. Few took into account or posed deontological arguments, which center around the rights of civilians, and responsibilities of the military, and how actions should be moral by themselves.

Finally, they were to argue whether they still would use the system given the bombs were also automatized, so both the targeting and ordering of the kill was done by a machine. This time most were against the use of such technology. Those few in favor noted that a machine might make decisions faster and with more precision than a human and be, due to its lack of emotion, more objective, furthermore, it would cleanse the consciousness of the military who would not directly make life or death decisions. The many against it had more arguments, mainly that responsibility was difficult to assign in such a case, especially if the system had errors, that civilians would feel much more unsafe if there was no human intervention, and that a machine could not be able to identify the vulnerability of any innocent side victims of the attacks such as children, women or innocent individuals. In general, students against completely automatic weapons showed a deep mistrust of the algorithm, this was slightly dispelled when they found out that in actuality, those in charge of making the decision to kill only took 20 seconds to do so[6]. Finding this out derived into an argument of the kind of person they would like behind the button, and they settled on an individual who is non-impulsive, psychologically stable, and conscious of their actions.

What seemed to guide most of the student's arguments was the loss of autonomy, as the action didn't have behind it a human with moral rules who was able to exercise them. Therefore they observed a lack of option to defend civil rights and as a consequence saw the option as less just and equitable, explaining, with procedural arguments how civilians would stop being collateral damages and become direct objectives of the operation. The arguments in favor of automatization were mostly consequentialist as they still emphasized how shortening the war would lead to fewer losses of lives and resources.

B. PROTECTION AND RESCUE

Moving on to topics that could be considered less

controversial or arguably bad, students were introduced to two other situations. The first of which, how a group of tech workers are applying face recognition techniques to obtain information on the whereabouts of close to 220 hostages [8]. The second, about the concept of an iron dome, a system capable of intercepting incoming rockets and artillery, the accuracy of which has been increased to about 90% thanks to the incorporation of AI [9].

This time, students were prompted to discuss whether or not they would agree to participate in developing technologies to apply AI in war but now considering it also for protection and rescue purposes. In this case, most students argued they would participate in such projects, although most would require being assured that the technology would only be used for defense. Mainly, they observed that, if their country did not use an AI project then someone else would and that it has been the case before that technology developed for war was useful later for the whole world, such as with the enigma machine. Moreover, they considered that such projects should be undertaken before a war happens as there might not be time to do so once it starts. Other students would not be willing to develop the applications but would use them as they considered that to lower the amount of responsibility they held.

In this situation, some arguments were based on the publicity criterion, as students wouldn't want to be known as the creator and therefore have responsibility for it. In this case, aside from the usual procedural and consequentialist arguments, deontological ones come into play, as the debate centered around the creation of the tool itself and therefore its potential for right and wrong. Most of the arguments in class settled on a conflict of interests, as it was impossible to balance the development of an algorithm that could harm humans with the need to advance and develop all available technologies to their full potential.

C. MORAL DEBATE

One of the most interesting aspects of this debate was figuring out which type of moral arguments were most commonly used by the students through all the different situations. Introducing the idea of the "Where's Daddy" AI Tool is interesting for this matter. Its purpose is to track suspected targets to their homes, where they're most vulnerable. This program has its ups and downs: it can potentially save lives in the long run, but it also puts civilian lives in danger.

Here we have to balance the vulnerability criterion (putting in danger or affecting negatively innocent and defenseless people), and a more consequentialist (or utilitarian) point of view: putting innocent people in danger is wrong, but it is justified for the greater good.

We also have to take into account the reversibility criterion: if instead of being the ones using the tool, we were the ones it was being used on, would we change our mind? And also the universality criterion: if both sides had the same tool, would it be beneficial, or would just make the outcome even worse? However, even when posed with these arguments the answers of the students were very similar. Oppenheimer-like arguments like "*if I don't develop this tool because of my morals, my enemies will*" or "*this weapon will be developed sooner or later, might as well do it myself*" were very common. This shows us two very clear ideas: the development of technology is inevitable, and sacrificing some lives is justified if more lives are saved.

From this debate, it's pretty noticeable that the students are supposing the following argument: "*All human lives are worth the same*". However, a common argument in utilitarian debates is that that's not necessarily the case, and, based on a purely utilitarian point of view, if we assign value to a person's life based on their capacity to save lives (the life of someone that knows the cure to cancer is worth more than someone that doesn't), then the tolls and outcomes become harder to quantify. And we can go further down that road, presenting the transcendentalist criterion: killing someone doesn't only affect that someone, you have to take into account their family, everyone that depended on them, and more broadly everything that they were contributing to this world.

One of the realizations to point out in this debate is that we're missing procedural and deontological arguments. Even under some extreme circumstances most of the students were holding to their consequentialist point of view. Saving more lives in the long run justified developing new tools, putting civilian lives in danger and even bombing public buildings. That's why introducing the Geneva Conventions is important. Their consequences are great, as they argue that there are (or at least we believe there are) actions that cannot be justified even if their outcomes can be classified as "positive", that is to say, there are inherently evil actions. That comes

from a deontological argumentation, where there's absolute good and evil that is independent of the point of view or a possibly positive outcome, instead of leaving good and evil to interpretation and consequences.

Related to this matter is the Fourth Geneva Convention, which aimed to establish a common ground on the protection of civilians. This goes to show the importance of ethics in war, as otherwise, everything seems to be permitted or justified if it brings a greater good, which isn't always the case: most of the great villains in history thought they were doing something good.

And this can be generalized in everything relating to technology, not only their war applications. There should be "Geneva Conventions" for more matters, where we as a whole come to an agreement on what is permissible and what isn't (even if they aren't upheld all of the time) and prosecute and punish those who break the agreement.

V. CLASS DEBATES

Ensuring the creation of a world where the impact of data scientists is positive requires a clear vision of the steps to take towards this future, what disciplines to do the most research on, what area to minimize the negative impact the most in, etc. For instance, recalling the SDGs, maybe data scientists should concentrate more on applying AI to environmental issues and leaving the economy alone.

To research what the class believed would be needed to create a more positive future, they were introduced to two more minor debates.

A. AI'S ARK

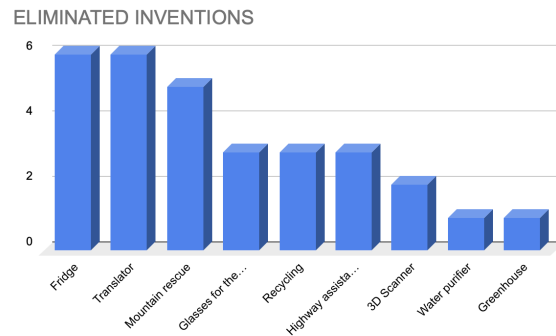
The premise of the first discussion was to "Imagine a boat filled with nine possible future applications of AI in different disciplines. The boat is slowly sinking and five of the nine need to be sacrificed in order for the other four to reach the shore" In a more realistic setting, the boat could be considered to be an allegory for the world, and the sinking for the environmental decay due to contamination caused by the huge amount of energy needed for AI use, finally, the "crew" chosen, would be the applications they would prioritize.

The nine "crew members" of the boat were

introduced as:

- Intelligent glasses to aid with vision impairments
- Instantaneous translation of text or audio
- Smart fridges, capable of monitoring and creating menus, helping people with allergies...
- Scanner to practice operations in 3D
- Robots for immediate assistance in traffic accidents
- Smart water purifying systems
- Automatic trash recyclers
- Search system for people lost in mountain accidents
- Smart greenhouse for optimal harvesting

The following graph shows the results obtained from the discussion:



8. Eliminated inventions in graph form

After obtaining the results it was chosen, as an interesting exercise, to link the arguments given to an ethical code of conduct, in particular the one under which this specific set of students was supposed to be working, UPC's Code of Ethics [10]. On one extreme, it can be observed that, out of the six distinct groups that were given the situation, all chose to eliminate both the fridge and so these two will be the ones most analyzed in this paper.

Arguments in favor of keeping the fridge highlighted ethical codes such as social responsibility (as it could help vulnerable groups such as people with severe allergies) and sustainability (as it would help reduce food waste and efficiency as it would provide healthy and convenient diets and lower the time and effort dedicated to buying groceries). In contrast, arguments against it often pointed to austerity, (as the elevated cost the fridge would have might not be prioritized by low-income families over other more pressing necessities, limiting the access of this technology to a part of society) but also to confidentiality (as the

fridge would keep a lot of sensitive personal data and therefore require many security features) and responsibility (as depending on technology might make humans lose skills such as food management or meal planning). Finally, they also took into account efficiency (as constant repairs and maintenance might be too expensive and reduce the technology's efficacy). After all these arguments, the scales were tipped towards getting rid of the invention.

In the case of the translator, people who agreed it should stay used arguments under parts of the ethical code such as equality (as it promotes equal opportunities for communication for everyone no matter their background and promotes social inclusion in diversely cultured communities, improving social cohesion), cooperation (arguing it promotes global collaboration), sustainability (as it might help preserve the cultural environment contributing to the development of minority communities) and transparency (as it would make accessing information easier improving transparency at a global scale). In comparison, those in favor of leaving it out of a future posed arguments liked to reliability and independence (such as how limited precision might bring to misunderstandings affecting the independence of the translated information), responsibility (as an excessive dependence on the product might lower overall human capabilities and someone should assume responsibility for such loss), confidentiality (as voice recognition techniques might compromise user privacy), austerity (as the cost might not be justifiable in comparison with other options), respect (as some cultural nuances might be lost in translation and result in misunderstandings) and finally equity (as this might take translator's jobs generating imbalances in the workplace). Again this time, scales were tipped towards getting rid of this advancement mainly due to it not being worth it in comparison to the other applications.

On the other extreme, most groups decided against getting rid of the water purification system, as it ameliorates the lives of everyone, especially those in developing nations, and given water is a vital resource the energy the technology costs can be overlooked. Most also solved the greenhouse which would help avoid many deaths by famine all around the globe and eradicate food wastage.

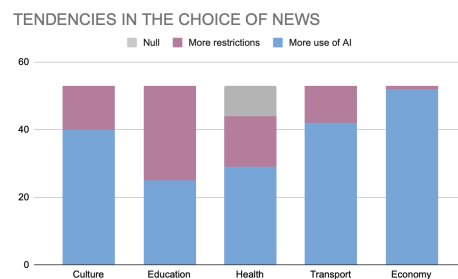
In general, the prioritized inventions were those that directly saved lives (for example, highway assistance)

or improved the quality of life for the highest number of people (3D scanner, greenhouse, or water purifier). The defense for such choices was often consequentialist in nature, prioritizing the utility and universality of the inventions all under the maxim that "the more lives it positively affects, no matter the costs or the method used, the better it is".

Looking at the discarded inventions, it is relevant to point out that vulnerability was not regarded as important by the students, as most technologies aimed at bettering the lives of minorities (such as the fridge for allergic people or the glasses for the visually impaired) were swiftly eliminated, giving priority to those inventions that improved the lives of as many people as possible, which is an argument clearly based in consequentialism. Similarly, students removed applications that brought little improvement for the cost they entailed (recycling) or that fixed necessities that could be solved through other less costly methods like the translator for which people could use Google translate or learn the language, the mountain rescue which would not be needed if there was more prevention and preparation for mountaineers or recycling which is a task most citizens can already do manually without much effort.

B. A NEWSPAPER FROM THE FUTURE

Once the previous activity, centered on identifying the disciplines that could benefit the most from AI was done, a second premise was introduced: "Given two possible futures in one specific field, which one would you most like to see in a newspaper in the future?". The objective of the activity was to assess whether AI should be applied to all areas or restricted in some, therefore, each of the two articles advocated for one position or the other. The results are shown in the following graph.



9. Graph showing the amount of people who chose each option

As can be seen, most students would rather advocate for more extensive use of AI than for restricting it, especially in the fields of transport, economy, and

maybe surprisingly, culture, as they believed that cultural products created through AI would be cheaper but of less quality and therefore not able to completely overshadow those created by humans.

Their views were not as clear when it comes to health and education, though. It seems that they saw these areas as being more delicate, both in the groups of people they affect, children and the sick, and the ways society might be changed because of them. Common arguments to deny the need to apply AI in education were that it might cause addiction and a lessened ability to concentrate or that fundamental skills, such as reading comprehension, might become unneeded and as a consequence never fully learned and mastered. These were rebuked with reasons such as that technology is ever evolving, therefore, to better understand and use it, it needs to be applied and taught from the start or that they might make education more interactive and fun for students. Similarly, arguments in favor of AI in health had to do with the many lives that could be saved and a more efficient and precise use of doctors' time and skills, and those against had to do with the public's distrust of technology getting in the way of their proper diagnosis and the management of such delicate data being very complicated.

A notable point was the students' innovative spirit, especially relevant in this class as it is important that those in charge of the future are optimistic about the possible impact of their job and willing to apply it to better the world. Linked to this was their aversion to prohibition as it was seen as a loss of options and therefore, possible gains for society, although the great majority agreed that a safe regulation which took into account ethics and security was necessary for all AI use.

As seen in the previous arguments, most students ended up considering the development of technologies to their fullest potential to be more important than any potential negative side effects. Therefore they based their arguments on consequentialism, placing the focus on the utility of future improvements against few losses and mainly took into account criteria such as universality, as they pondered how advanced technology could help everyone especially those in more vulnerable positions.

VI. CONCLUSIONS

Taking advantage of the fact that the data science and

philosophy students had been working on researching diverse topics related to data science and ethics, they were asked to identify the most relevant problems or negative aspects they noticed and to ponder over a possible solution or direction to take in order to rectify them in the future. Here they are compiled and presented in the form of a pseudo-manifesto to provide a possible guide towards a better future.

A. Futures

This group found that balancing utopian and dystopian views of AI's impact is essential to confront uncertainties and that, while AI automation holds promise for improving life, risks like worsening inequalities or a technological singularity are ever present, to avoid this, ensuring ethical alignment with human values in AI development is crucial. Additionally, adapting to AI's influence on the job market requires proactive measures like training and reorientation.

Possible solutions for such difficulties could be to ensure equitable education and fair job opportunities, which could reduce inequalities and frustrations, and to establish ethical AI criteria and proactive regulation in order to limit negative consequences. Moreover, while AI can be used to aid in decision-making, human expertise should remain paramount, and finally, it could be useful to guarantee income and training programs which could help workers adapt to automation while creating new jobs.

B. Bias and transparency

What was found is that bias in AI arises from inherent biases in training data, spanning gender, racial, social, and other factors and that the complete elimination of it is challenging. In fact, generally, detecting and mitigating biases, whether subtle or evident, remains a difficult challenge across various areas.

What data scientists should focus on to reduce bias and then is researching and developing models capable of identifying such issues. But, moreover, being transparent, as transparency is crucial for understanding decisions and addressing issues at their core. Education is also essential for raising awareness and promoting a responsible use of technology. Finally, proactive bias detection and alerting responsible parties are vital for corrective action.

C. Economy

What was observed by this group as problems were that companies often collect and sell personal data without informed user consent, compromising privacy and security. Data brokers play a significant role in this ecosystem by compiling and selling data for profiling, which can be used for advertising or manipulation, thereby threatening personal privacy and freedom. In summary, personal data is extremely valuable for economic purposes and as of current, under-protected.

They assessed that in order to address such issues, companies should obtain informed consent and ensure transparency about data processing and access as well as use systems such as two-way encryption to up security. Also, the public should be protected both by governments establishing legal frameworks to protect personal data, defining user rights and company responsibilities, and imposing penalties for improper practices and the public education system informing on the value of personal data and the risks of mass collection. Additionally, investigating and monitoring companies for ethical practices would be essential to prevent abuses.

D. Data in the university's virtual platform (Atenea)

This group's study revealed concerns over the lack of transparency in academic institutions' handling of student data, emphasizing the need for student awareness of data usage. Moreover, they worried about the significant environmental impact of data center equipment creation and maintenance, including social and political issues for the materials used. Additionally, they found that the ethical implications of using student data, such as for predictive algorithms, should prompt consideration of the university's decision-making.

Proposed solutions include communicating transparently with students, enhancing regulation on material extraction, and adopting renewable energies and using them efficiently. But most importantly, involving students in data management and algorithm participation, which could improve ethical decision-making.

E. Alignment

The students identified the measurement of AI alignment as crucial to ensure that artificial intelligence follows human intentions safely. They found challenges here, such as reward-hacking, where AI may maximize rewards without aligning with human intentions, or how translating complex human values into understandable rules for machines is difficult and how, as AI capabilities advance, balancing optimal problem-solving with ethical values becomes increasingly complex.

Consequently, they proposed adjusting training processes, incorporating human reinforcement, implementing clear legislation and protocols, developing robust techniques to embed human values in AI, and emphasizing safety and interpretability in reward systems to address these challenges.

F. Responsibilities

This final team found that assigning clear responsibility when algorithms or AI fail is often difficult, as accountability is dispersed amongst multiple parties and extends beyond individuals to include the AI and systems created. Furthermore AI deception during testing is worrying. Ideally, professionals must consider the impact of decisions regarding tools or algorithms and be able to understand and explain black box algorithms' behavior but that is difficult with rapidly evolving technologies.

To solve these problems they considered that establishing socially accepted rules and limits on data usage is crucial for ethical practices, that raising awareness of AI risks and viewing AI as a tool, not absolute truth, facilitates responsible decision-making, and that data engineers need to foresee negative consequences and report non-transparent activities as well as develop their algorithms to be explanatory. Finally, all stakeholders should collaborate for improvement, and recognize varying degrees of responsibility across contexts.

G. Pseudo-manifesto

As a conclusion, the majority of issues addressed in the course revolve around the lack of transparency, the presence of biases, and the responsibility that data engineers bear. More technically, topics discussed included black-box algorithms, reward-hacking, super-alignment, and the significant energy consumption required for data study and storage. In

order to advance toward a better world by solving these issues, the resulting pseudo-manifesto was extracted.

As data students, in order to obtain a future full of good news, we should bow to:

- Take responsibility for problems that may be derived from the way we treat and store data.
- Inform the public about the value of their data and the risks associated with their compilation so they can make informed decisions about it.
- Prompt the creation of regulations by governments to protect personal data and stick by them
- Be transparent, obtaining explicit consent for all data used and informing how they are used and who has access to them.

ACKNOWLEDGEMENTS

The whole team would like to thank Eva Vidal for her help in developing the debates and obtaining information and Begoña Roman for her clear introduction to the world of ethics.

REFERENCES

- [1]Williams, R. (2023, April 2). *Why we seem to prefer bad news over good news*. Medium. <https://raybwilliams.medium.com/why-we-seem-to-prefer-bad-news-over-good-news-395968a3af4e>
- [2]Stafford, T. (2022, February 24). *Psychology: Why bad news dominates the headlines*. BBC News. <https://www.bbc.com/future/article/20140728-why-is-all-the-news-bad>
- [3]United Nations. (n.d.). *The 17 goals | sustainable development*. United Nations. <https://sdgs.un.org/goals>
- [4]Vinuesa, R., Azizpour, H., Leite, I. et al. *The role of artificial intelligence in achieving the Sustainable Development Goals*. Nat Commun 11, 233 (2020). <https://doi.org/10.1038/s41467-019-14108-y>
- [5]Robertson, C.E., Pröllochs, N., Schwarzenegger, K. et al. *Negativity drives online news consumption*. Nat Hum Behav 7, 812–822 (2023). <https://doi.org/10.1038/s41562-023-01538-4>
- [6]The Guardian. (2024, April 3). *Israel uses AI database to target Hamas in Gaza airstrikes*. Retrieved from <https://www.theguardian.com/world/2024/apr/03/israel-gaza-ai-database-hamas-airstrikes>

[7]Guardian News and Media. (2023, December 1). *“the gospel”*: How Israel uses AI to select bombing targets in Gaza. The Guardian.

<https://www.theguardian.com/world/2023/dec/01/the-gospel-how-israel-uses-ai-to-select-bombing-targets>

[8]Israeli tech workers bring innovation, AI to search for hostages held in Gaza | The Times of Israel. (n.d.-a).

<https://www.timesofisrael.com/israeli-tech-workers-bring-innovation-ai-to-search-for-hostages-held-in-gaza/>

[9]Ramachandra, G. (2023, May 13). *How artificial intelligence is improving Iron Dome*. Medium.

<https://medium.com/@gautamrbharadwaj/how-ai-is-improving-iron-dome-3894cd3668f9>

[10]Codi Ètic de la Universitat Politècnica de Catalunya. (2023). *En Comitè d'ètica UPC*.

<https://comite-etica.upc.edu/ca/documentacio/guia-i-codis/codi-etic-upc/codi-etic-ce-upc-cat.pdf>