ISSN 1870-4069

AI Safety in Mexico: A Pilot Survey in Yucatan

Janeth Valdivia Pérez[1,3], Valeria Ramirez Hernández[1], Silvia Fernández-Sabido[2,3], Ángel Tenorio Vázquez[3], Alejandro Molina-Villegas[2], Oscar Sánchez Sordia[2]

[1] Universidad Politécnica de Yucatán, Mexico

[2] Centro de Investigación en Ciencias de Información Geoespacial, Mexico

[3] AI Safety Mexico Project, Mexico

{janeval92,valery.ramirez.hdez}@gmail.com, {sfernandez,osanchez}@centrogeo.edu.mx, aetenorio@gmail.com

Abstract. As artificial intelligence transforms society, understanding regional perspectives on its risks becomes vital. This study presents a pilot survey conducted in Yucatan, Mexico, aimed at capturing local concerns about AI safety. Inspired by the global survey *"Thousands of AI Authors on the Future of AI"*, this initiative was a joint effort of the Universidad Politécnica de Yucatán and the Centro de Investigación en Ciencias de Información Geoespacial, in Merida, Mexico. The survey is piloted on a small sample of academics, including students, professors, and researchers, probing their views on the ethical, safety, and regulatory aspects. The survey delved into topics such as the explainability of AI, its potential risks, and the implications of its advancements over the next twenty years. These preliminary findings reveal concerns about the spread of false information, manipulation of public opinion, and the possibility of AI being used in an authoritarian manner. These concerns are consistent with global trends and reflect unique characteristics rooted in the socioeconomic and cultural realities of Mexico. Notably, 97% of respondents agreed that AI experts should consider the needs and concerns of society, while 83% believe the government should invest in mitigating AI safety risks. Furthermore, 81% expressed support for the creation of a special AI safety agency in Mexico. This kind of study contributes to the academic and political dialogue about AI safety in Mexico and also lays the groundwork for future research and the development of policies that promote a safe and ethical implementation of AI.

Keywords: AI safety, survey analysis, ethical implications, regulation, public perception.

1 Introduction

In recent years, artificial intelligence (AI) has experienced exponential growth, revolutionizing industrial, academic, and social sectors globally. Like many other countries, Mexico is at a crucial juncture where the adoption and integration of AI in various areas present unprecedented opportunities as well as significant challenges. Given this dynamism, AI safety has become an area of growing interest among the academic and scientific community, emphasizing its importance in promoting responsible and ethical technological development.

In this context, the implementation of a pilot survey aims primarily to evaluate the effectiveness of selected questions to measure key concerns of the study population. The questions focus on aspects such as the explainability of AI, the causes of its progress, potential risks over the next 20 years, and general perceptions of its safety in Mexico. Moreover, this pilot survey seeks to refine the methodologies used to ensure the accuracy and relevance of the data collected. This preliminary effort is crucial for preparing a more extensive national survey focused on AI safety in Mexico.

The study, inspired by the ESPAI (AI Impacts' Survey on AI Progress) 2023 [1], adapts and selects relevant questions for the Mexican context and introduces new inquiries focused on local concerns about the use and regulation of AI.

The momentum for this survey comes from a short (two month) professional stay collaboration between the Centro de Investigación en Ciencias de Información Geoespacial (CentroGeo) and the Universidad Politécnica de Yucatan (UPY), in Merida Mexico, where resources from both institutions have been combined to deepen understanding of the current AI landscape in the region and to ensure that the design and implementation of the survey provide a clear and detailed vision that will guide future initiatives and policies in this emerging field. Through this study, we aim to establish a solid foundation for future research and policies that guide the safe and ethical development of artificial intelligence in the country.

2 Objective

The main goal of the study is to conduct a pilot survey on AI safety within the academic community (students, professors and researchers), with the purpose of capturing the perceptions and concerns of the study population. This will serve as preparation for a more comprehensive and detailed national survey on the same topic.

The specific objectives are to translate and select questions from the ESPAI 2023 survey conducted by the AI Impacts team that may be useful to understand the perspective on AI safety in Merida, Yucatán and; to create a block of questions to help explain the perception of the use and regulation of AI safety in Mexico. The survey also aims to refine the methodologies employed to ensure the accuracy and relevance of the data collected.

3 Related Work

Artificial intelligence (AI) safety research in Mexico has evolved, albeit at a slower pace compared to the global scene. Challenges in regulation and safety span a range of concerns from ethical design to the mitigation of unintended consequences of AI systems. This research often intersects with broader discussions on data protection, human rights, and socioeconomic impacts, facilitated by interdisciplinary teams composed of scientists, philosophers, and policymakers.

Initiatives like the *Instituto Nacional de Acceso a la Información y Protección de Datos Personales* (INAI) highlight the need for privacy and data ethics in AI, setting a precedent for incorporating ethical considerations in AI development [5]. Meanwhile, the academic community in Mexico has begun to address the issue of bias in AI systems. Studies like those by Ramirez emphasize that factors such as limited access, insufficient resources to evaluate data, and inherent discrimination in AI systems can deepen existing gaps in digital, social, political, and economic realms. They also raise ethical and moral questions about the implications of the development of current AI systems [7].

In terms of regulation, in 2018, the Mexican government, in collaboration with Oxford Insights, C-Minds, the Mexican Society for Artificial Intelligence, and *Tecnológico de Monterrey*, supported financially by the UK Embassy in Mexico, embarked on a mission to develop specific actions the government could take to promote the development and use of AI across all sectors of the country. However, with the change of government.

In the same year, the planned AI agenda was not continued. Nevertheless, entities like the UNAM Institute of Legal Research worked in their Public Policy Reports during 2021 to identify challenges and obstacles for designing a public AI policy that included a Human Rights approach in the country, emphasizing the importance of creating a National Artificial Intelligence Strategy with diverse perspectives to identify potential risks in AI development and deployment [9].

Since 2019, Mexico has committed to adhering to the OECD principles for responsible AI development by signing the first set of intergovernmental policy guidelines on AI, agreeing to comply with international standards that ensure the design of AI systems are robust, secure, fair, and reliable [3].

Recently, during 2023 and 2024, deputies and senators from various political parties in Mexico have presented and discussed important initiatives in the Chamber of Deputies and the Senate to pass laws on regulation to establish measures that ensure the transparency and oversight of AI systems, however, to date, no initiative has been approved yet [8,2,6].

As we can observe, research on AI safety in Mexico is in a stage of growth and alignment with international standards, facing challenges including ethics, privacy, and bias in algorithms. Collaboration among academia, government, and international organizations is crucial to creating a robust framework that ensures the responsible development of AI. Recent legislative initiatives in Mexico reflect a growing commitment to the regulation and supervision of technology.

4 Methodology

4.1 Survey Themes

With the aim of achieving the goals of this study, a survey titled "Survey on AI Progress" was created, aimed at capturing the perception of students, professors, and scientists primarily from Merida in Yucatán Mexico. The survey was inspired by the AI Impacts Expert Survey on AI Progress 2023 (ESPAI 2023) [1] and the article "Thousands of AI Authors on the Future of AI" developed by the AI Impacts team [4,1].

ESPAI 2023 is a key survey that measures the perceptions of AI researchers, following the methodology of its predecessors, the ESPAI surveys of 2022 and 2016. With 2,778 responses from researchers, this survey represents a significant sample of the AI academic community.

The questions focused on consistent themes over the years to compare the evolution of opinions and included some new ones, designed after an iterative testing process. After multiple revisions of the survey and article, a total of 44 questions were formulated that assess the topics shown in Table 1. By adopting this proven and detailed methodology, our survey aims to effectively capture opinions and attitudes towards artificial intelligence among the academic community, and reinforce the communication work on the topic of safety in Mexico in order to detect specific challenges and opportunities.

Table 1. Key AI Topics: Development and Impact.

AI Development	**Safety & Society**
The explosion of intelligence	20-year AI traits
Explainability	Tasks (fixed probabilities and fixed years)
Scenarios	Safety quote
Value	Safety resources + ethics
Causes of AI Progress	Safety resources
HLMI outside view	Extinction
Meta and sociology	Demographics
High-level machine intelligence (HLMI), through jobs (fixed probabilities and fixed years)	Speed and safety

4.2 Survey Design and Data Collection

The importance of basing our survey on the ESPAI questions to capture perceptions about AI safety in Mexico lies in ensuring relevance and accuracy. In this sense, we adhere to the wording of the original questions, seeking a careful

translation into Spanish and selecting those we consider of interest for the current study. It was also considered prudent to retain the Likert scale proposed in the original survey and to develop a block of questions according to our local context. The survey included the sections shown in Table 2.

Table 2. Sections of the survey

Research Areas	Safety Considerations
Collection of demographic data	Safety
High-Level AI and Total Job Automation	Safety resources and ethics
AI Intelligence Explosion	Perception of AI safety in Mexico
Explainability	Feedback section
Causes of AI Progress	Completion time
AI risks within 20 years	

We selected a sample of 36 people, 42% of whom are students from the Universidad Politécnica de Yucatán, studying Data Engineering, Cybersecurity, and Robotics at undergraduate and graduate levels. The remaining 58% are researchers and/or professors from institutions such as the Centro de Investigación y de Estudios Avanzados del IPN (Cinvestav), the Centro de Investigación Científica de Yucatán (CICY), and the Instituto de Investigaciones en Matemáticas Aplicadas y Sistemas (IIMAS), dedicated to fields like data science, software development, and machine learning, among others.

The survey was created in Google Forms, distributed via invitation link through email and WhatsApp. Due to the short duration of the stay and because it was a pilot test (we left the bulk of the community for the future survey), it was launched on April 1, 2024, and remained open until April 8 of the same year. At the end of the survey period, we obtained a CSV file of semi-structured data without defined data types or relationships. In this regard, we developed a data dictionary to assist in our preprocessing and transformation code.

We focused on data extraction and preprocessing in Colab, a hosted Jupyter notebook service. It is also important to note that, given the small sample size, it wasn't advisable to stratify it, but in the future we consider differentiating between students, professors, and researchers, and even between those who have had contact with the topic of AI-Safety and those who haven't.

4.3 Data Cleaning

During the development of a Python script designed for analyzing a survey on the Progress of Artificial Intelligence, we utilized several highly recognized libraries (Pandas, Numpy and Matplotlib) in the field of data science, which facilitated both the manipulation and visualization of data. The analysis began with an exploratory examination of the data.

To facilitate handling during the analysis, we opted to rename the columns using a 'Q' followed by a number format (e.g., 'Q1', 'Q2', etc.). This change not only simplified referencing the columns but also improved the organization of the DataFrame.

Before proceeding with data deletion, a careful selection of questions deemed optimal for the study was made. These questions included topics on explainability, causes of AI progress, potential AI risks in 20 years, and the perception of AI Safety in Mexico. After defining the key questions for this analysis, several columns that were not necessary were removed, thus reducing the dataset's dimensionality and focusing the analysis on the most relevant variables for the study.

4.4 Data Analysis and Visualization

The analysis of the data collected from the Survey on the Progress of Artificial Intelligence was deepened through the generation of descriptive statistics and the use of effective visualization techniques. All data were analyzed based on percentages, a crucial practice in this type of study primarily because it allows for uniform comparison between groups of different sizes. In surveys and polls, converting raw counts into percentages normalizes the data, which is useful for performing statistical analysis and for visualizing trends more effectively, especially when dealing with data categorized across multiple levels or scales.

Descriptive statistics were calculated for each of the variables of interest, this approach helped to identify preliminary patterns, understand the distribution of responses, and establish a baseline for more detailed analyses.

4.5 Visualization Techniques

Visualization techniques used to further explore the data included bart charts, stacked bar charts, and heat maps, which allowed for comparing frequencies and intensity of responses in percentage terms across different question categories, highlighting significant differences and similarities. Finally, the results of the analysis will be interpreted in the context of the study's objectives. Conclusions and recommendations will be drafted aimed at promoting effective safety and regulation of AI in Mexico.

5 Development

5.1 Concern Levels for AI Scenarios over the Next 30 Years

Through the question: "How would you rate the level of concern that the following possible AI-related scenarios deserve over the next 30 years?", eleven AI scenarios that could be a cause for concern for society were assessed. In Figure 1 we can observe how the respondents evaluated the severity of the following scenarios for the next 30 years.

Of the scenarios presented, those that generated the most concern when adding up the percentages of the most representative categories of response such as "Quite worrying" and "Extremely worrying" were the spread of false information, like deep fakes which had a cumulative 97.2% of substantial or extreme concern, followed by the manipulation of public opinion on a large scale representing 88.9% while concern about Authoritarian rulers using AI to control their populations represented 83.3%.

Other categories such as scenarios like AI systems that worsen economic inequality by disproportionately benefiting certain individuals and the possibility of AI allowing dangerous groups to create powerful tools, like designed viruses represented 80.6% and 77.8% respectively.

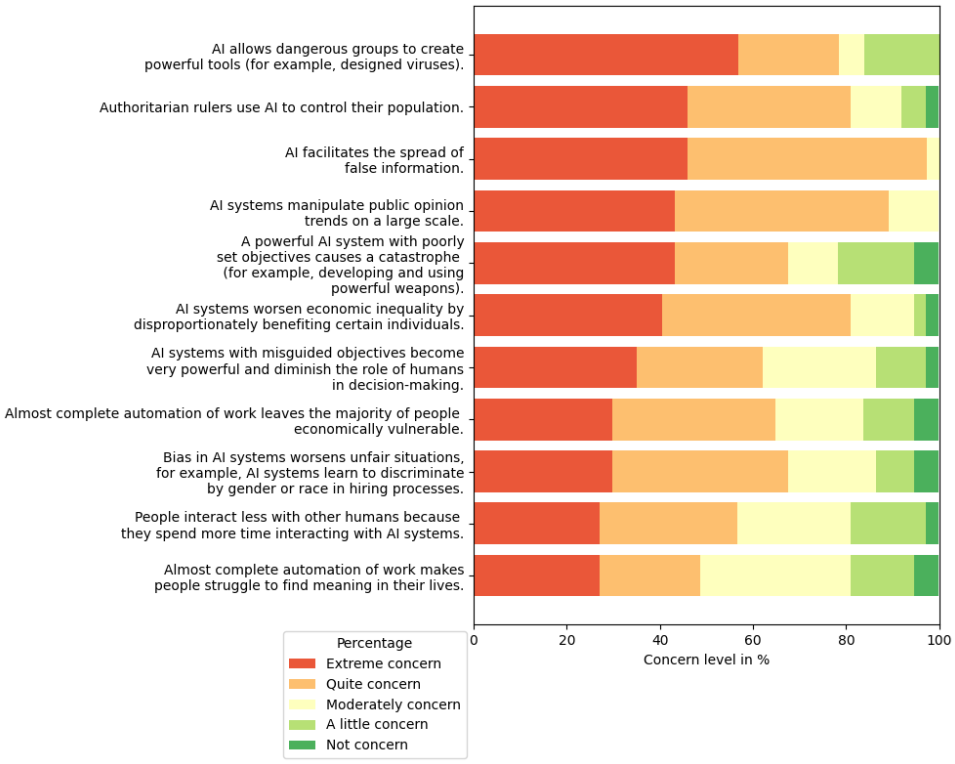

Fig. 1. Concern levels for AI scenarios over the next 30 years. Source: own elaboration with data obtained from the Survey on the progress of AI,

5.2 Causes of AI Progress

The analysis of the data collected on the causes of AI progress (Figure **??**) revealed that overall, the factors of research effort, computing power, quantity and quality of data, progress in AI algorithms and funding have been extremely important as causes of AI progress. However, it is notable that more than 60% of the participants classified computing power as the main cause, while the quantity and quality of data was considered the second most representative cause with 58.3%. It also reveals that 55.6% of participants consider research effort to be extremely important, while funding and progress in algorithms stand out with 41.7% and 44.4% participation, respectively.

Some of the scenarios, such as "Finding unexpected ways to achieve goals" which represents a 44.4% probability, and "Making design improvements to increase their own performance regardless of human desires" at 38.9%, have a high perception of being "Extremely likely". Other scenarios, such as "Can speak as a human expert on most topics" and "Often behaving in ways that are surprising to humans", also have a considerable perception of probability.

Figure 3 shows descriptive statistics of the five categories assessing the perception of risk associated with artificial intelligence (AI) over a span of 20 years, according to respondents' answers. It is observed that the mean progressively increases from "Not at all likely" to "Extremely likely," reflecting a rise in the perceived probability across categories. Concurrently, the standard deviation shows an upward trend, while the median (50%) maintains consistency with the mean, rising with each subsequent category.

As for extreme values, the maximum observation is recorded in the "Extremely likely" category at 44.40%, followed by 38.90% in the categories of "Slightly likely" and "Quite likely". The minimum value (min), in turn, rises from 0 in "Not at all likely" to 16.70 in "Equally likely", and again to 23.60% in "Extremely likely".

Finally, it is noteworthy that the gap between the third quartile (75%) and the maximum value is not particularly wide in the "Quite likely" and "Extremely likely" categories. This phenomenon is also illustrated in Figure 2, where the values tend to cluster towards the upper end of the scale in these categories.

6 Perception of AI Safety in Mexico

In Figure 4 we analyze perceptions about the impact and regulation of artificial intelligence (AI) in Mexico. The results are presented in percentages reflecting the respondent answers to various statements:

- *Labor market and equity concerns*
  Regarding the statement "AI experts should take into account the needs and concerns of society", the graph shows an almost unanimous consensus among participants (98%) on the importance of AI experts considering society's needs and concerns, as they agreed or strongly agreed with this statement.

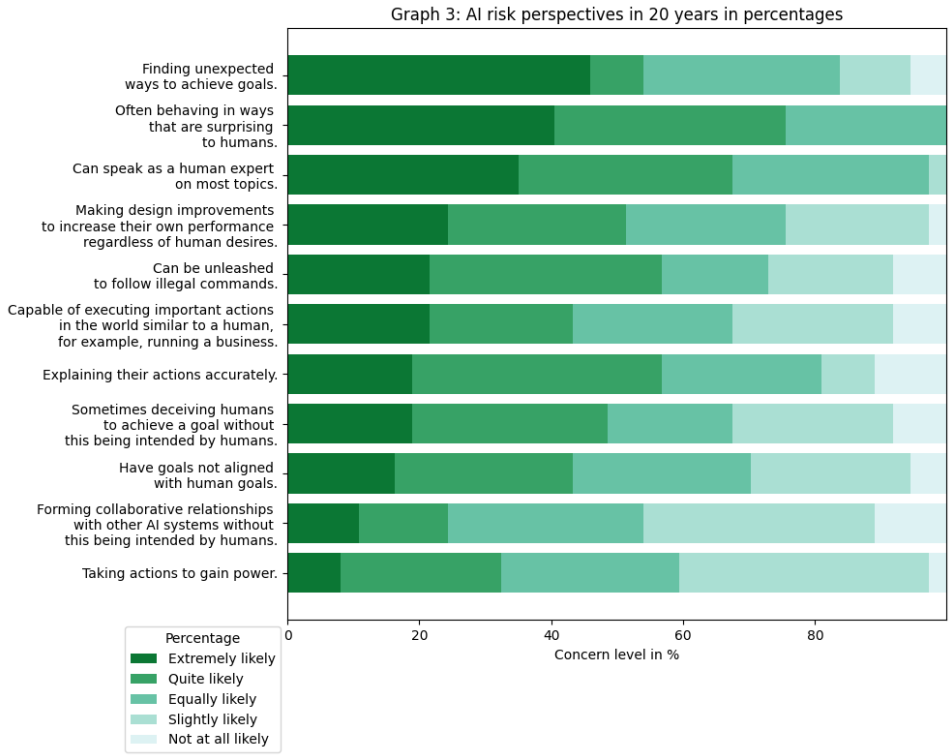

Fig. 2. AI risk in 20 years. Source: own elaboration with data obtained from the Survey on the progress of AI.

	Not at all likely	Slightly likely	Equally likely	Quite likely	Extremely likely
count	11.00	11.00	11.00	11.00	11.00
mean	5.05	18.67	25.77	26.77	23.72
std	3.88	12.11	4.51	9.41	11.48
min	0.00	0.00	16.70	8.30	8.30
25%	1.40	9.70	25.00	23.60	18.05
50%	5.60	22.20	25.00	27.80	19.40
75%	8.30	23.60	29.20	33.35	30.55
max	11.10	38.90	30.60	38.90	44.40

Fig. 3. Descriptive statistics of the expressed probabilities of the risk of AI at 20 years. Source: own elaboration with data obtained from the Survey on the progress of AI.

Concern about the impact of AI on the labor market in Mexico is significant, with 66.6% of participants agreeing or strongly agreeing with this concern. Similarly, there is a notable worry about biases in AI algorithms and the risk they pose to equity and justice in Mexico, with 61.1% expressing agreement or strong agreement.

– *Government regulation and supervision*
  A significant majority of respondents believe in the government's importance in regulating and supervising AI, with 75% agreeing or strongly agreeing that the government should regulate and supervise AI more.
– *Investment in safety and specialized agency*
  The government's investment in mitigating AI-related safety risks is considered important, with 83% agreeing or strongly agreeing on its significance. Furthermore, 81% of respondents would agree with the creation of a special agency for AI safety in Mexico.
– *Perception of current regulations*
  Regarding the adequacy of Mexico's current regulations to meet AI safety challenges, only 11% of participants agreed or strongly agreed that they are adequate — suggesting general skepticism about the current regulatory framework.
– *Socioeconomic concerns*
  Finally, the possibility that socioeconomic inequality in Mexico could increase due to the way AI is currently being developed and utilized in the country is a concern for 61% of participants, who agreed or strongly agreed with this statement.

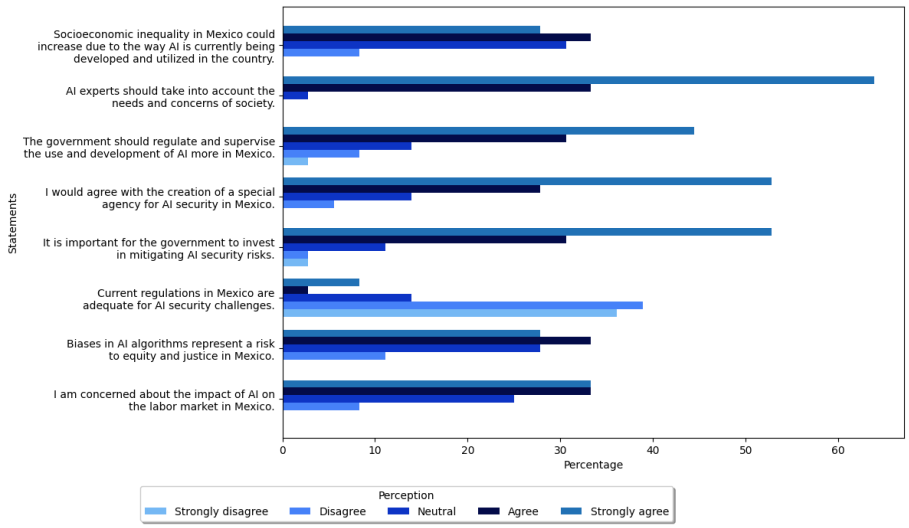

Fig. 4. AI safety perceptions in Mexico. Source: own elaboration with data obtained from the Survey on the progress of AI[10].

7 Discussion

Comparing analysis of the survey results from the ESPAI 2023 as outlined in the document "Thousands of AI Authors on the Future of AI" and those derived from the "Survey on AI Progress" tailored for Merida, Yucatan, we delve into the nuanced perceptions and concerns surrounding the safety and advancement of artificial intelligence (AI). Highlighting both contrasts and alignments in global and local contexts. The section on AI risk scenarios in the local study indicated a high level of concern for scenarios such as the spread of false information and the large-scale manipulation of public opinion by AI systems. There was also significant worry about the use of AI by authoritarian rulers to control populations. The findings published in "Thousands of AI Authors on the Future of AI" similarly reflected significant concerns over similar scenarios, such as the spread of misinformation and authoritarian control over populations. However, a distinctive aspect was the emphasis on the exacerbation of economic inequality noted in the ESPAI 2023.

Analysis of the causes of AI progress showed that the main perceived drivers include computing power, data quantity and quality, and research efforts. These results align closely with those found in the original survey, with the addition of AI algorithmic progress as a significant driver. Regarding AI risk perspectives over the next 20 years, responses in the survey indicated that scenarios such as "Finding unexpected ways to achieve goals" and "Making design improvements to enhance their own performance, regardless of human desires," were seen as extremely likely.

On the perception of AI safety in Mexico, there was significant consensus on the importance of government regulation and oversight of AI. A high percentage of respondents supported the idea of creating a specialized AI safety agency in Mexico. Additionally, there was considerable concern about the potential impact of AI on the labor market and the biases in AI algorithms that could affect equity and justice in Mexico.

As for expectations about AI progress, the ESPAI 2023 survey anticipated significant advancements, with a considerable proportion of respondents foreseeing the possibility that AI could outperform humans in all possible tasks in the not-too-distant future. However, the focus of the adapted survey was more on understanding perceptions of AI-related risk scenarios and their social, ethical, and safety implications.

While the original survey did not delve deeply into specific regulations, it suggested a general need for research on AI safety and addressing potential risks of the technology.

It is important to note that creating a section to analyze perceptions of AI safety in Mexico showed clear consensus on the importance of government regulation in overseeing AI. This idea is reinforced by a high percentage of respondents supporting the creation of a specialized agency for AI safety in Mexico.

Overall, while the original survey provided a more global view focused on the pace of AI technological advances, the adapted survey for Merida concentrated

more on understanding local perceptions of AI risks and regulation, reflecting specific concerns about safety and ethics in the use of technology. Both surveys, however, underscore the importance of addressing the risks associated with AI and the need for effective regulation to ensure responsible and safe technological development.

Given that the Python script was designed to complete the first step of data extraction in an ETL (Extract, Transform, Load) process and includes various transformation operations such as data cleaning, handling null values, renaming columns to facilitate analysis, and some basic mathematical and statistical operations that are part of the transformation process, it is believed that the future can improve and complete an ETL process for handling survey data at a national level.

It is also believed that expanding the survey's scope to include more regions of Mexico could provide a broader view of the perception of AI safety throughout the country.

8 Conclusion and Future Work

The conclusion of this report highlights how the survey on AI safety in Merida Yucatán, adapted from the study "Thousands of AI Authors on the Future of AI", provides valuable insights into the perception of risks, advancements, and regulations of artificial intelligence in a local context. Through comparative analysis, it was evident that there is high concern for scenarios such as the spread of false information and the manipulation of public opinion by AI systems, similar to concerns observed in the global context. However, a distinctive aspect was the emphasis on the need for AI experts to consider the needs and concerns of society, demonstrating the specific interests of the local environment.

Moreover, the analysis of the causes of AI progress revealed that research efforts, computing power, and data quality are seen as key drivers of AI advancement, closely paralleling those found in the original survey. This underscores the consistency in the perception of AI technological progress at both local and global levels.

The survey also showed strong consensus on the importance of government regulation and supervision of AI, with significant support for the creation of a specialized AI safety agency in Mexico. This finding is crucial as it highlights the need for an informed governance strategy specifically directed at managing AI risks while leveraging its potential benefits.

Regarding the performance and scope of the project, the methods and tools used for data collection and analysis proved effective in capturing and understanding complex perceptions of AI safety. However, areas of opportunity were identified to improve and complete the ETL process, which could further optimize data management and analysis in the future. Expanding the survey's focus to include more regions of Mexico could enrich our understanding of how AI is viewed and regulated across different cultural and economic contexts within the country.

We also considered stratifying the sample to capture differences between students, professors, and researchers, and even between those who have had contact with the topic of AI-Safety and those who have not.

This report not only contributes to the academic and political dialogue on AI safety in Mexico but also provides a solid foundation for future research aiming to contribute to the generation of knowledge in this area to promote the design of responsible AI technologies ethically aligned with the social and cultural values of Mexico and the world.

Acknowledgments. The first author would like to thank Eng. Angel Tenorio for his participation as the project manager of the team, to Valeria Ramirez, a data engineering student, who participated in the adaptation of the survey, as well as Alex Molina and Oscar Sánchez for his feedback during the development of this project. But most importantly, to Silvia Fernandez for her invitation to participate in this project and her support as an external supervisor. This study was partially funded by Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI) and Centro de Investigación en Ciencias de Información Geoespacial (CentroGEO), from Mexico.

References

1. AI Impacts: About ai impacts. `https://aiimpacts.org/about/` (2023), last accessed: 2024-04-03
2. [Anonymous]: Sesgos y discriminaciones sociales de los algoritmos en inteligencia artificial. Entretextos, (2024)
3. Cortés, V., Ruíz, P.: Reportes de política pública. Report, Jurídicas UNAM (2024), `https://archivos.juridicas.unam.mx/www/site/index/-8330.pdf`, Último acceso: 2024-04-03
4. Gentzkow, K.: Thousands of ai authors on the future of ai. arXiv, (2024)
5. Instituto Nacional de Transparencia, Acceso a la Información y Protección de Datos Personales (INAI): México: Inai recommendations on ai. `https://www.dataguidance.com/opinion/mexico-inai-recommendations-ai` (2024), Último acceso: 2024-04-03
6. Monreal Ávila, R.: Iniciativa con proyecto de decreto por el que se expide la ley federal que regula la inteligencia artificial (2024), `https://ricardomonrealavila.com/wp-content/uploads/2024/02/Inic_Morena_inteligencia_artificial.pdf`, Último acceso: 2024-04-03
7. Morales Ramírez, G.: Problemática antropológica detrás de la discriminación generada a partir de los algoritmos de la inteligencia artificial. MYE, vol. 34, no. 2, pp. 429–480 (2023)
8. Organisation for Economic Co-operation and Development (OECD): Oecd legal instruments. `https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449` (2024), last accessed: 2024-04-03
9. Presidencia de la República EPN: Inteligencia artificial en México. `https://www.gob.mx/epn/es/articulos/inteligencia-artificial-en-mexico` (2024), Último acceso: 2024-04-03
10. Pérez, J. V.: Survey on AI safety. `https://github.com/JanneVa/Survey_on_AI_Safety` (2025), gitHub repository. Last accessed: 2025-05-05