THE IMPACT OF DATA PROCESSING ON THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE IN AFRICAN COUNTRIES

By Cecil Abungu

Strathmore University

Centre for Intellectual Property and Information Technology Law

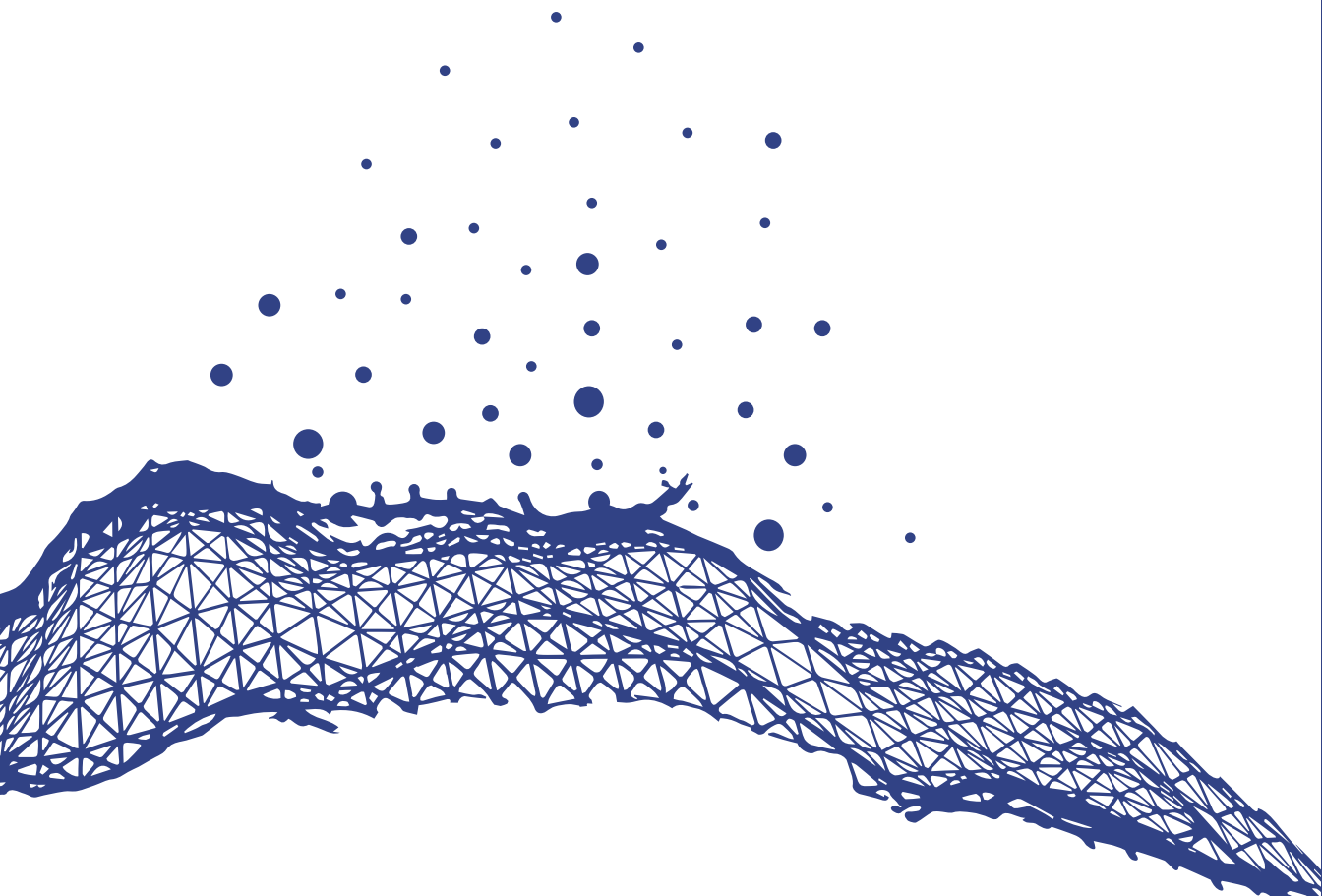
Contents

Table of Figures

Tables

INTRODUCTION

The main aim of this research is to examine what impact the principles of data minimization and data non-repurposing are having or are likely to have on the training of algorithms in African countries. This is with an objective to recommend the best way to ensure a balance between committedly protecting individual human rights while at the same time building a competitive AI ecosystem.



DEMOGRAPHICS

The study sought to determine the characteristics of the target population to understand the general population. Some of this information included the African country where the respondents worked, the respondents' field of work and their main are of focus in artificial intelligence (AI).

Place of work

Figure 1 shows a representation of the countries where the respondents work.

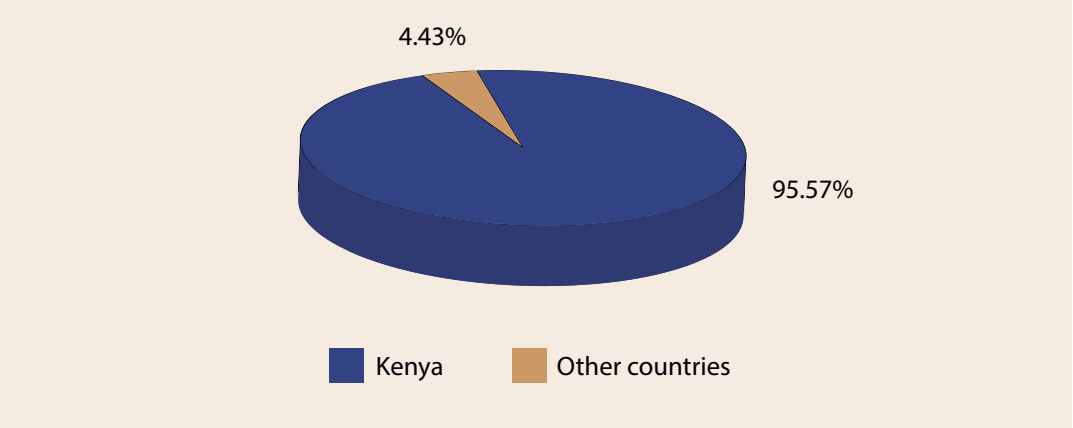


Figure 1: Place of work

The study reached 158 respondents with most of the respondents (151) working in Kenya, contributing to 95.57% of the total respondents while the remaining 4.43% (7) coming from other African countries.

Current field of work/current area of research

The study sought to establish the main fields of work or research of the respondents and the data is shown on figure 2.

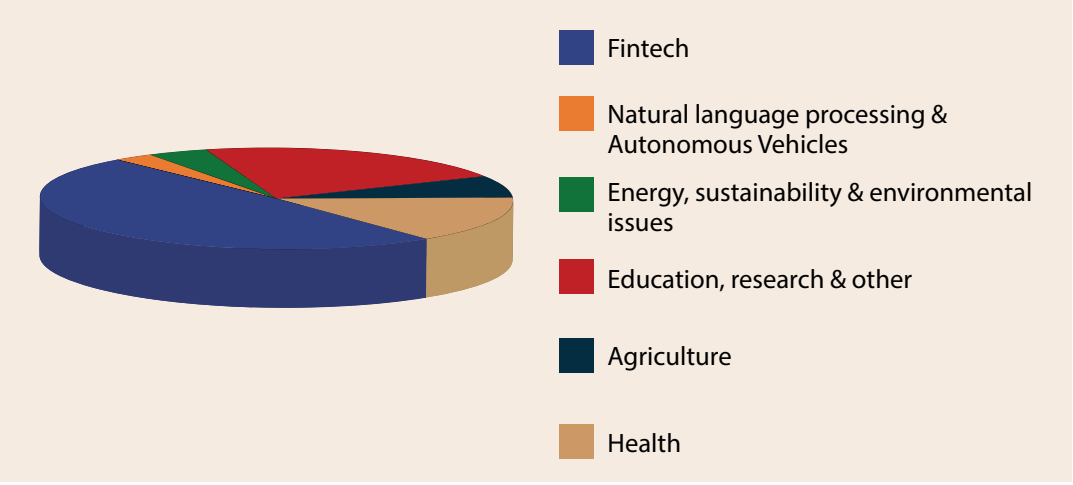


Figure 2: Field of work/research

Most respondents (49%) indicated that they worked or did research in the fintech field followed by those in education and research. The least number of respondents (5) said they worked in the Natural Language Processing & Autonomous Vehicles field, representing 3% of the total respondents.

Current area of focus in Artificial Intelligence

Machine learning and data mining and analysis were the two current areas that most respondents said they currently focussed in as represented in figure 3.

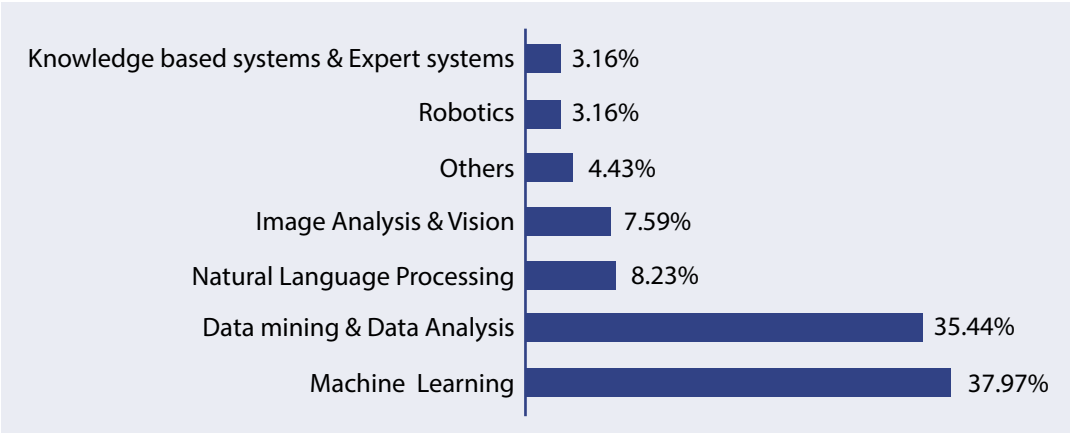


Figure 3: Current area of focus in AI

These two areas contributed to 73.41% of the total respondents, with machine learning having more respondents focusing on it at 37.97%. This was followed by the Natural Language Processing area with 8.23% of the respondents. Knowledge Based Systems & Expert systems and Robotics fields had the least respondents focusing on them with 3.16% of the respondents, each.

Sources of data in the respondents' current field of specialization

One of the objectives of this research was to establish the sources of data for the respondents in their current fields of specialization and results are shown in figure 4.

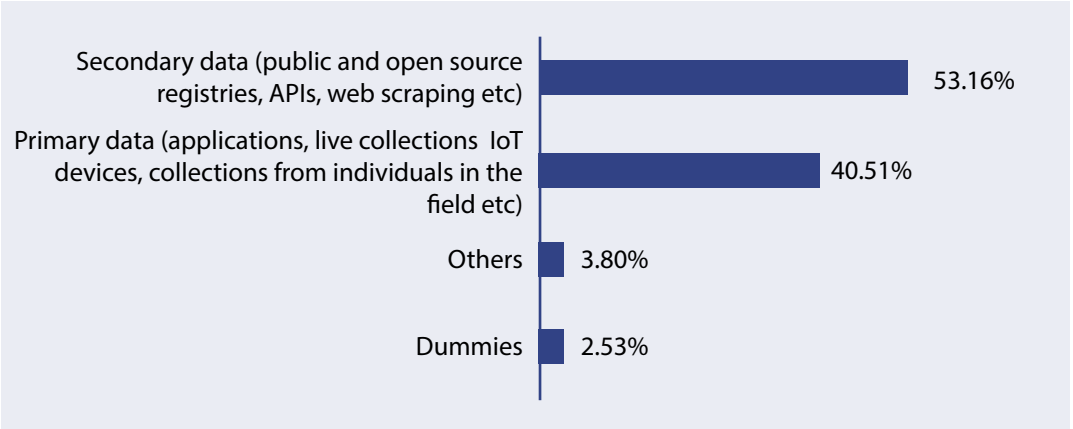


Figure 4: Sources of data

Findings show that secondary data from public and open-source registries, APIs, etc., was the main source of data for most respondents contributing to 53.16%. Primary data came in second as a source of data for 40.51% of the respondents while use of dummy data was a source of data for the least number of respondents at 2.53%. Other data sources accounted for 3.85 of the respondents.

Criticality of data in respective fields of AI

Most respondents affirmed that data is very critical in the field of AI that they are in, as shown in figure 5.

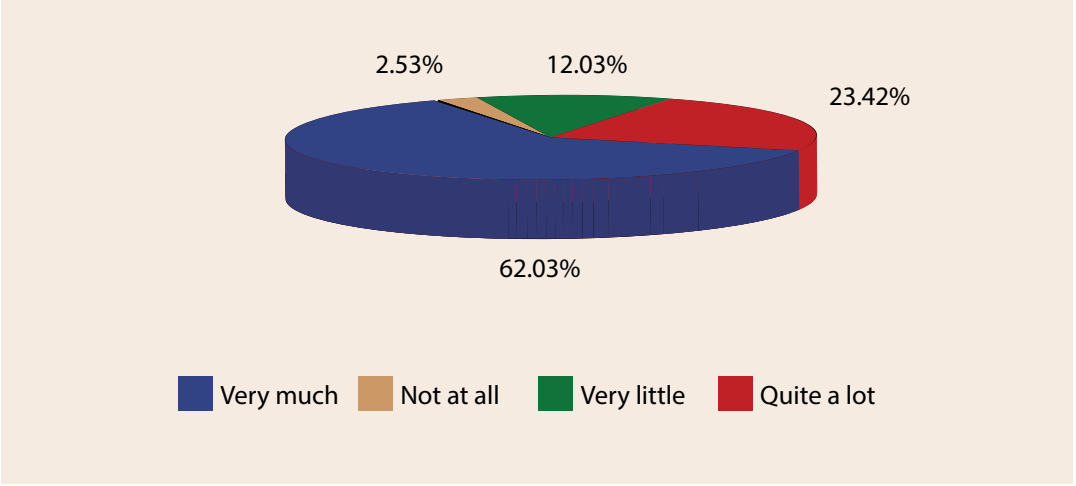


Figure 5:Criticality of data in the field of AI

85.45% of the respondents (62.03% saying it was very much critical and 23.42% saying it was critical quite a lot) said data was a critical issue for them in their fields with only 2.53% of the respondents saying that data was not a critical issue at all in their respective fields of AI.

DATA ANALYSIS AND FINDINGS

Data Minimization

The study sought to establish whether respondents were aware of the principle of data minimization, regulations around this principle and its impact in their respective fields of AI. Data minimization in this study implies processing of data that are adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.

Awareness of regulations on data minimization

On the awareness of regulations on data minimization in the fields of AI of the respondents and in the countries where they worked, only 32.91% of the respondents indicated that they were very aware of those regulations as shown in figure 6.

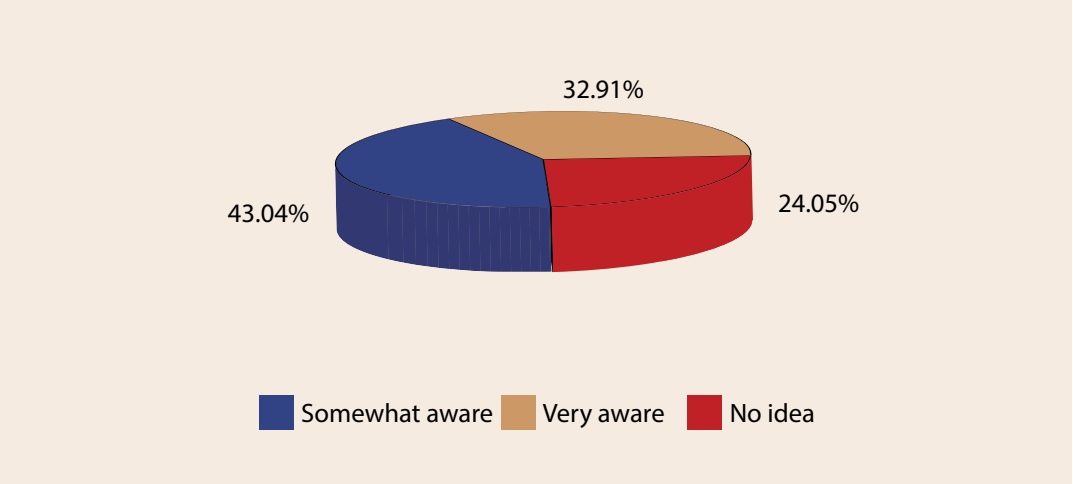


Figure 6: Awareness of regulations on data minimization

Majority of the respondents (43.04%) said they were somewhat aware of the regulations while the remaining 24.05% had no idea whatsoever on data minimization regulations in their fields of AI. Of those who were aware of the regulations, most of them (24) said they have made t difficult for them to train AI models, followed by those who did not see a difference (21) in the effect of these regulations on their ability to train AI models. The remaining respondents indicated that the regulations made it easy for them to train AI models (5) and others did not know of the impact of these regulations (2) as shown in table 1.

Data minimization in this study implies processing of data that are adequate, relevant, and limited to what is necessary in relation to the purposes for which they are processed.

Table 1: Awareness of data minimization regulations and their impact on ability to train AI models

How aware are you of the existence of such regulations?					Total
How have these regulations affected your ability to train AI models?		No idea	Somewhat aware	Very aware	
I do not know	Count	13	3	2	18
	% within	34.2%	4.4%	3.8%	11.4%
They have made it difficult	Count	9	36	24	69
	% within	23.7%	52.9%	46.2%	43.7%
No difference	Count	15	27	21	63
	% within	39.5%	39.7%	40.4%	39.9%
They have made it easy	Count	1	2	5	8
	% within	2.6%	2.9%	9.6%	5.1%
Total	Count	38	68	52	158
	% within	100.0%	100.0%	100.0%	100.0%

A majority of respondents who had no idea of data minimization regulations (15) said they saw no difference in the effect of those regulations on their ability to train AI models. These were followed by those who did not know whether the regulations had any impact on their training AI models (13). Those who found it difficult to train AI models but did not know of these regulations came in third (9) while the least number of respondents (1) who had no idea of the regulations found it easy to train AI models.

Most respondents indicated that inasmuch as they had an idea of these regulations, they did not understand the scope. For instance, some did not know what data bulkiness entailed. They assumed that one needs to collect as much data as they could for better insights. Those who were not aware of these regulations were willing to learn about them and how they affected the AI industry. For respondents from Kenya who were aware of the regulations, they indicated that the regulations were limiting and as such were engaging stakeholders to have an AI policy, which is under development, ratified soon. The lack of a policy has hindered many projects and limited AI developers. They also suggested information literacy trainings for persons to ensure compliance and awareness of data processing regulations.

Impact of data minimization regulations on ability to train AI models

Most respondents (43.68%) thought that regulations on data minimization have made it difficult for them to train AI models as shown in figure 7 (31.65% said it is difficult and 12.03% indicating the regulations have made it very difficult).

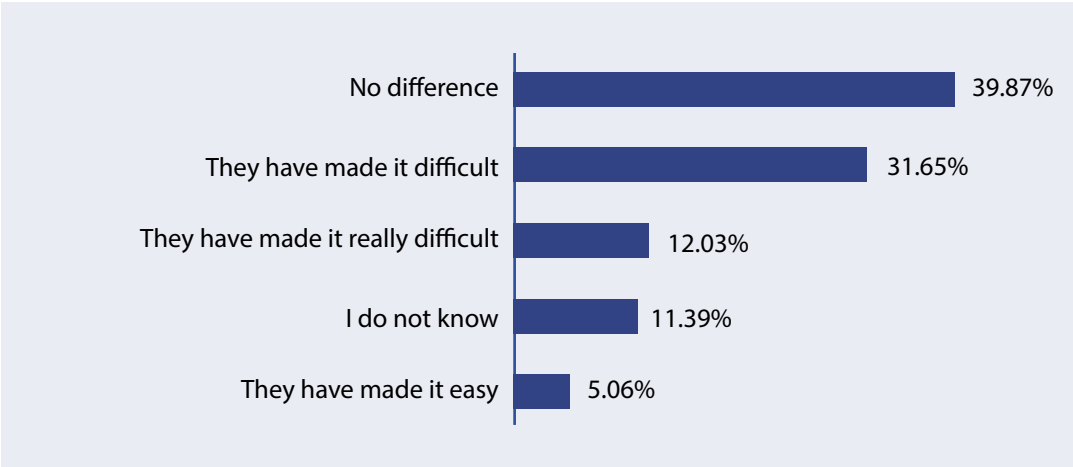


Figure 7: Impact of data minimization regulations on ability to train AI models

This was followed by 39.87% of the respondents who felt that the regulations have had no impact on their ability to train AI models with 11.39% not knowing the impact these regulations have had on their ability to train AI models in their fields. The remaining 5.06% of the respondents said the regulations on data minimization in their fields of AI made it easier for them to train AI models.

Going by their current field of work, most respondents (33) who said that data minimization regulations made it difficult for them to train AI models came from the Fintech field as shown in Table 2. These were followed by those in Education, research, and others with 19 respondents and those from the health field came third (8).

Table 2: Impact of data minimization regulations on ability to train AI models per respondents' field of work

	How have these regulations affected your ability to train AI models?					Total
Current field of work		I do not know	They have made it difficult	No difference	They have made it easy	
Agriculture	Count	5	2	3	1	11
	% within	45.5%	18.2%	27.3%	9.1%	100.0%
Natural Language Processing & Autonomous Vehicles	Count	1	2	2	0	5
	% within	20.0%	40.0%	40.0%	0.0%	100.0%
Fintech	Count	7	33	35	4	79
	% Within	8.9%	41.8%	44.3%	5.1%	100.0%
Health	Count	2	8	12	2	24
	% Within	8.3%	33.3%	50.0%	8.3%	100.0%
Energy, sustainability & environmental issues	Count	0	5	1	0	6
	% Within	0.0%	83.3%	16.7%	0.0%	100.0%
Education, Research & other	Count	3	19	10	1	33
	% Within	9.1%	57.6%	30.3%	3.0%	100.0%
Total	Count	18	69	63	8	158
	% Within	11.4%	43.7%	39.9%	5.1%	100.0%

Similarly, most respondents (35) who did not see a difference in the impact of data minimization regulations on their ability to train AI models were from the Fintech field followed by those in health (12) and then education, research, and others (10). This trend was the same for those who indicated that the regulations made it easy for them to train their AI models with a majority (4) being in the Fintech field, followed by health (2) and thirdly, education, research, and others (1).

Given the respondents' area of focus in AI, a majority of (28) respondents who indicated that data minimization regulations made it difficult for them to train their AI models were from the Data Mining and Data Analysis area followed by those in machine learning (21) and the third category were from Natural Language and Processing (8) as shown in table 3.

Table 3: Impact of data minimization regulations on ability to train AI models per respondents' area of focus in AI

	How have these regulations affected your ability to train AI models?					Total
Current area of focus in AI		I do not know	They have made it difficult	No difference	They have made it easy	
Robotics	Count	0	2	2	1	5
	% Within	0.0%	40.0%	40.0%	20.0%	100.0%
Machine learning	Count	10	21	26	3	60
	% Within	16.7%	35.0%	43.3%	5.0%	100.0%
Natural Language Processing	Count	1	8	4	0	13
	% Within	7.7%	61.5%	30.8%	0.0%	100.0%
Knowledge Based Systems & Expert Systems	Count	1	2	1	1	5
	% Within	20.0%	40.0%	20.0%	20.0%	100.0%
Image analysis and vision	Count	1	6	4	1	12
	% Within	8.3%	50.0%	33.3%	8.3%	100.0%
Data Mining and Data Analysis	Count	4	28	22	2	56
	% Within	7.1%	50.0%	39.3%	3.6%	100.0%
Others	Count	1	2	4	0	7
	% within	14.3%	28.6%	57.1%	0.0%	100.0%
Total	Count	18	69	63	8	158
	% within	11.4%	43.7%	39.9%	5.1%	100.0%

Respondents from the other areas of focus in AI who indicated that data minimization regulations made it difficult for them to train their AI models were from Image analysis and vision (6), Robotics (2) and others (2). Another significant number of respondents said that they did not think there would be any impact of data minimization regulations on the ability of respondents to train their AI models. In this category, most respondents were in machine learning (26) followed by data mining and data analysis (22). Natural language processing, image analysis and vision and others followed with 4 respondents each, robotics (2) and the least number of respondents were in Knowledge Based Systems & Expert Systems (1). The study noted that 5.1% (8) of the respondents indicated that the data minimization regulations made it easy for them to train AI models in their areas of expertise. Machine learning had 3 respondents, data mining and data analysis followed (2) while robotics, image analysis and vision, and others had 1 respondent respectively, with the same view.

Respondents (42) who got their data from secondary sources indicated that data minimization regulations have made it difficult for them to train AI models. These were followed by those whose source was primary data (23) and finally dummies (3) and other sources (1) as shown on table 4. Out of the 5.1% (8) who said that data minimization regulations made it easy for them to train AI models, 2.53% (4) indicated that they got their data from secondary sources, 1.9% (3) from primary data and 0.63% (1) from other sources.

Table 4: Impact of data minimization regulations on ability to train AI models by respondents' sources of data

	How have these regulations affected your ability to train AI models?					Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Primary data	Count	7	23	31	3	64
	% within	10.9%	35.9%	48.4%	4.7%	100.0%
Secondary data	Count	10	42	28	4	84
	% within	11.9%	50.0%	33.3%	4.8%	100.0%
Dummies	Count	0	3	1	0	4
	% within	0.0%	75.0%	25.0%	0.0%	100.0%
Others	Count	1	1	3	1	6
	% within	16.7%	16.7%	50.0%	16.7%	100.0%
Total	Count	18	69	63	8	158
	% within	11.4%	43.7%	39.9%	5.1%	100.0%

For those who indicated that they saw no difference in the effect of the data minimization regulations on their ability to train AI models, a majority (31) said their source of data was primary, followed by those whose source was secondary data (28). Those who used dummies and other sources od data were 1 and 3 respectively.

In addition to sources of data, the study sought to establish how data minimization regulations affected the respondents by their perspective of criticality of data in their field of AI. Table 5 outlines the effect of these regulations. Most respondents (62) who thought data was highly critical in their field of AI said that the regulations made it difficult for them to train AI models, they were followed by those who said data was not critical in their field but thought that the regulations made it difficult for them to train AI models (7).

Most respondents (54) who saw no difference on the impact of data minimization regulations on

their ability to train AI models said data was very critical in their area of AI while the least number (9) thought that data was not critical.

Table 5: Impact of data minimization regulations on ability to train AI models by criticality of data

Do you consider access to data to be a critical issue in the field of AI which you are in?		How have these regulations affected your ability to train AI models?				Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Not at all	Count	1	1	2	0	4
	% within	25.0%	25.0%	50.0%	0.0%	100.0%
Very little	Count	3	6	7	3	19
	% within	15.8%	31.6%	36.8%	15.8%	100.0%
Quite a lot	Count	1	20	14	2	37
	% within	2.7%	54.1%	37.8%	5.4%	100.0%
Very much	Count	13	42	40	3	98
	% within	13.3%	42.9%	40.8%	3.1%	100.0%
Total	Count	18	69	63	8	158
	% within	11.4%	43.7%	39.9%	5.1%	100.0%

For those who indicated that the regulations made it easy for them to train AI models, a majority (5) indicated that data was very critical in their area while the rest (3) said that data was not critical.

Respondents further claimed that these regulations have made persons from whom data needs to be collected malicious and would like to be paid to give data hence making the data collection exercise expensive and difficult. It is also worthy to note that the respondents found that policies around data minimization are not clear as to the scope hence some AI developers got more data from other countries as they are not limited to train their models. However, some felt that these regulations would limit their ability to have more data for training models.

Effect of data minimization regulations on ability of other AI developers to train AI systems

The study sought to know the perspective of respondents on the impact the data minimization regulations have on other AI developers that the respondents know in their respective countries.

The results are outlined on figure 8.

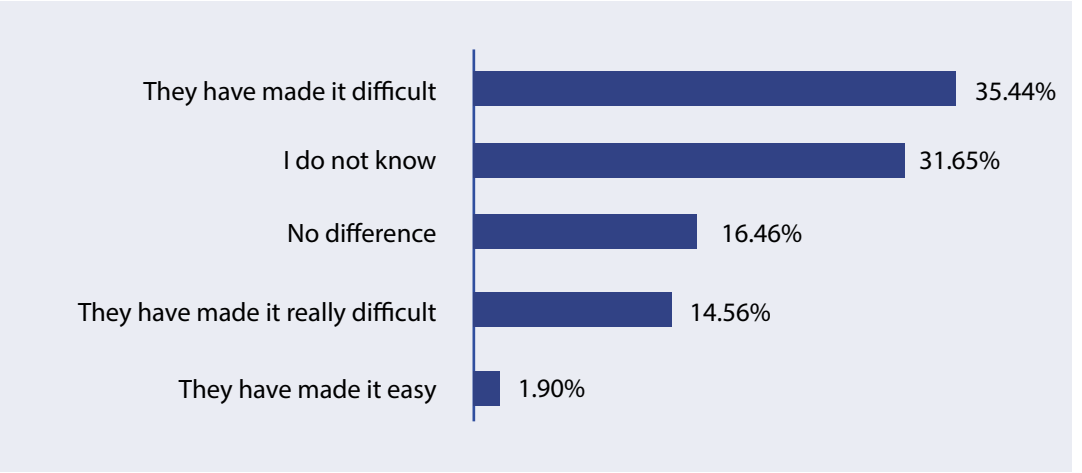


Figure 8: Effect of data minimization regulations on ability of other AI developers

Half of the respondents said that the data minimization regulations have made it difficult for other AI developers they know in their countries of work to train AI models (35.44% said the regulations have made it difficult and 14.56% said they have made it very difficult). Respondents who did not know the effect data minimization regulations had on other AI developers' training of AI models contributed to 31.65% of the respondents while those who thought there was no difference in the impact of the regulations accounted for 16.46%. the remaining 1.95 of the respondents thought that the data minimisation regulations made it easier for other AI developers to train AI models. Respondents further detailed that the difficultness is compounded by AI developers not defining the problem properly hence having to collect obsolete data as the problem keeps evolving over the development phase. Some respondents indicated that for them to find relevant and accurate data, one had to pay for it making it quite expensive and difficult to train models.

Impact of data minimization regulations on ability to train AI systems in future if unchanged

The study sought to establish what the respondents thought would be the impact of data minimization regulations on their ability to train AI models in the future if the regulations as they currently are in their countries. The findings are represented in figure 9.

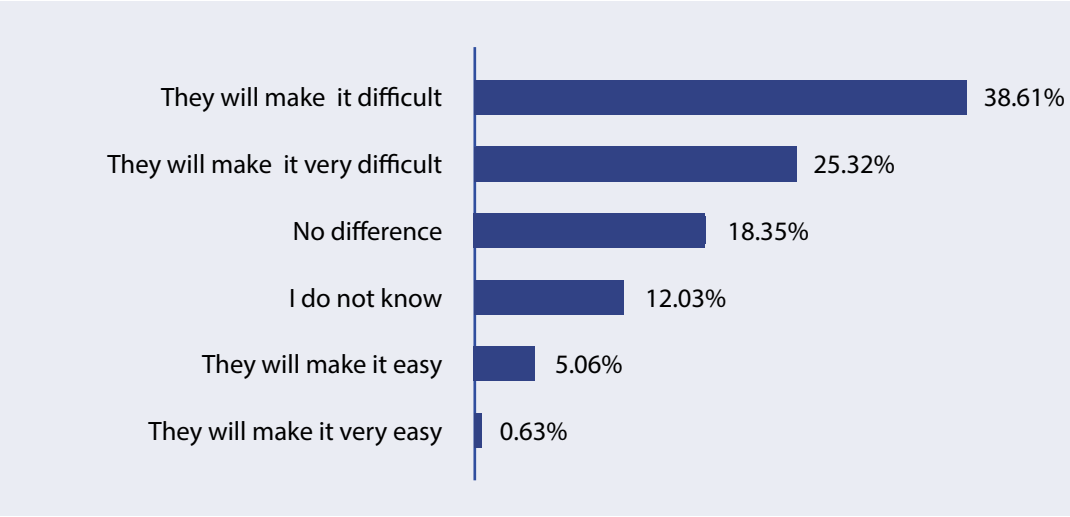


Figure 9: Impact of data minimization regulations on ability to train AI systems in future if unchanged

Most of the respondents (63.93%) said that if the regulations remained as they are, they would make training of AI models difficult in the future (38.61% thought they would make it difficult and 25.32% said it would be very difficult). Respondents who thought there would be no difference in impact if the regulations remained as they are accounted for 18.35% of the respondents while those who had did not know what the impact would be accounted for 12.03% of the respondents. The remaining respondents (5.69%) said the regulations would make training of AI models easy if they remained unchanged in the future.

Moreover, the study wanted to establish how respondents in their various fields of work thought data minimization regulations would affect their ability to train AI models if the regulations remained unchanged. There was an increase in respondents who said that if the regulations remain unchanged, it will make it difficult to train AI models. This is outlined in table 6. Slightly over half of these respondents (52) were from the Fintech field, followed by Education, research & others (21) and then health (15). The agriculture, natural language and processing and Energy, sustainability & environmental fields had the least respondents with 6, 4 and 3 respondents respectively.

Table 6: Impact of data minimization regulations on respondents' ability to train AI models in the future if regulations are unchanged

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
Current field of work		I do not know	They have made it difficult	No difference	They have made it easy	
Agriculture	Count	4	6	0	1	11
	% within	36.4%	54.5%	0.0%	9.1%	100.0%
Natural Language Processing & Autonomous Vehicles	Count	0	4	1	0	5
	% within	0.0%	80.0%	20.0%	0.0%	100.0%
Fintech	Count	8	52	15	4	79
	% within	10.1%	65.8%	19.0%	5.1%	100.0%
Health	Count	1	15	7	1	24
	% within	4.2%	62.5%	29.2%	4.2%	100.0%
Energy, sustainability & environmental issues	Count	0	3	2	1	6
	% within	0.0%	50.0%	33.3%	16.7%	100.0%
Education, Research & other	Count	6	21	4	2	33
	% within	18.2%	63.6%	12.1%	6.1%	100.0%
Total	Count	19	101	29	9	158
	% within	12.0%	63.9%	18.4%	5.7%	100.0%

Those respondents who indicated that there would be no impact on their ability to train models if these regulations remained unchanged dropped to 29 with the majority being from the Fintech field followed by health. Those who said that it will be easy for them to train AI models despite the data minimization regulation remaining the same in the future stood at 9, a negligible increase by 1 respondent. This was the same for respondents who said they did not know what impact the unchanged regulations would have on their ability to train AI models in the future.

Given the respondents' area of focus in AI, the study established that most respondents (45) in data mining and data analysis said that if the current data minimization regulations remain unchanged, it will be difficult to train AI models in the future. These we followed by respondents in machine learning (35) and image analysis and vision and Natural language processing with 9 respondents each indicating the same view of difficulty as shown on table 7.

Table 7: Impact of data minimization regulations if they remain unchanged on respondents' ability to train AI models in the future per

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
		I do not know	They will make it difficult	No difference	They will make it easy	
Robotics	Count	0	3	1	1	5
	% within	0.0%	60.0%	20.0%	20.0%	100.0%
Machine learning	Count	9	35	13	3	60
	% within	15.0%	58.3%	21.7%	5.0%	100.0%
Natural Language Processing	Count	0	9	4	0	13
	% within	0.0%	69.2%	30.8%	0.0%	100.0%
Knowledge Based Systems & Expert Systems	Count	2	1	1	1	5
	% within	40.0%	20.0%	20.0%	20.0%	100.0%
Image analysis and vision	Count	3	9	0	0	12
	% within	25.0%	75.0%	0.0%	0.0%	100.0%
Data Mining and Data Analysis	Count	3	41	8	4	56
	% within	5.4%	73.2%	14.3%	7.1%	100.0%
Others	Count	2	3	2	0	7
	% within	28.6%	42.9%	28.6%	0.0%	100.0%
Total	Count	19	101	29	9	158
	% within	12.0%	63.9%	18.4%	5.7%	100.0%

Most of those who indicated that it will be easy to train AI models if the regulations remain unchanged in future (4) were from data mining and data analysis, followed by those in machine learning (3) while robotics and Knowledge Based Systems & Expert Systems had 1 respondent each.

Most respondents (51) who got their data from secondary sources indicated that if the data minimization regulations remained unchanged, it would make training of AI models difficult in the future. This was followed by those whose source was primary data (44) and those who used dummies and other sources at 3 respondents each as shown on table 8.

Table 8: Impact of data minimization regulations if they remain unchanged on respondents' ability to train AI models in the future by source of data

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
		I do not know	They will make it difficult	No difference	They will make it easy	
Primary data	Count	2	44	15	3	64
	% within	3.1%	68.8%	23.4%	4.7%	100.0%
Secondary data	Count	16	51	11	6	84
	% within	19.0%	60.7%	13.1%	7.1%	100.0%
Dummies	Count	0	3	1	0	4
	% within	0.0%	75.0%	25.0%	0.0%	100.0%
Others	Count	1	3	2	0	6
	% within	16.7%	50.0%	33.3%	0.0%	100.0%
Total	Count	19	101	29	9	158
	% within	12.0%	63.9%	18.4%	5.7%	100.0%

Most respondents (6) who said they thought it would be easy for them to train AI models in the future if data minimization regulations remained unchanged got their data from secondary sources while the rest (3) got it from primary sources.

The study further sought to establish the perspective of respondents on the impact of data minimization regulations on their ability to train AI models in the future if the regulations remain unchanged. Table 9 shows that most respondents (91) who indicated that data is very critical in their area will find it difficult to train AI models in future if the regulations remain unchanged. The rest (10) who found data to be not critical indicated that it would be difficult to train AI models in future if regulations remained unchanged.

Table 9: Impact of data minimization regulations if they remain unchanged on respondents' ability to train AI models in the future by data criticality

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
Do you consider access to data to be a critical issue in the field of AI which you are in?		I do not know	They have made it difficult	No difference	They have made it easy	
Not at all	Count	2	2	0	0	4
	% within	50.0%	50.0%	0.0%	0.0%	100.0%
Very little	Count	3	8	7	1	19
	% within	15.8%	42.1%	36.8%	5.3%	100.0%
Quite a lot	Count	4	23	7	3	37
	% within	10.8%	62.2%	18.9%	8.1%	100.0%
Very much	Count	10	68	15	5	98
	% within	10.2%	69.4%	15.3%	5.1%	100.0%
Total	Count	19	101	29	9	158
	% within	12.0%	63.9%	18.4%	5.7%	100.0%

Conversely, most respondents (8) who found data to be very critical in their area of focus said that it would be easy for them to train AI models in future if the regulations remained unchanged while the remaining (1) respondents who indicated that data was not critical in their area said it would also be easy to train models in future if the regulations remained unchanged.

The study noted that most respondents (29) who said they were aware of data minimization regulations indicated that if they remained unchanged, it will be difficult for them to train Ai models in the future as shown on table 10. These were followed by those (13) who said there would be no difference in the impact on training AI models if the regulations remained unchanged in the future. The remaining respondents (10) who were aware of data minimization regulations indicated that they would not know (5) the impact of the regulations or it would make it easy (5) for them to train AI models in the future if the regulations remained unchanged.

Table 10: Awareness of data minimization regulations and their impact on respondents' ability to train AI models in the future if they remain unchanged

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?					Total
		No idea	Somewhat aware	Very aware	
I do not know	Count	5	9	5	19
	% within	13.2%	13.2%	9.6%	12.0%
They will make it difficult	Count	27	45	29	101
	% within	71.1%	66.2%	55.8%	63.9%
No difference	Count	5	11	13	29
	% within	13.2%	16.2%	25.0%	18.4%
They will makeit easy	Count	1	3	5	9
	% within	2.6%	4.4%	9.6%	5.7%
Total	Count	38	68	52	158
	% within	100.0%	100.0%	100.0%	100.0%

For those who were not aware of the data minimization regulations, a majority (27) indicated that it would make it difficult for them to train AI models in the future if the regulations remined unchanged. These were followed by those who did not know the impact (5) and those who saw no difference (5) on the impact of these regulations if they remained unchanged in future. Those that indicated that it would be easy to train AI models despite the regulations remaining unchanged but had no idea of the regulations were the least (1) of these respondents.

Impact of data minimization regulations on ability of other AI developers to train AI systems in future if unchanged

On the impact of data minimization regulations on other AI developers' ability to train AI models in future if the regulations remained as they are, most respondents (63.3%) said they would make it difficult. This is shown on figure 10.

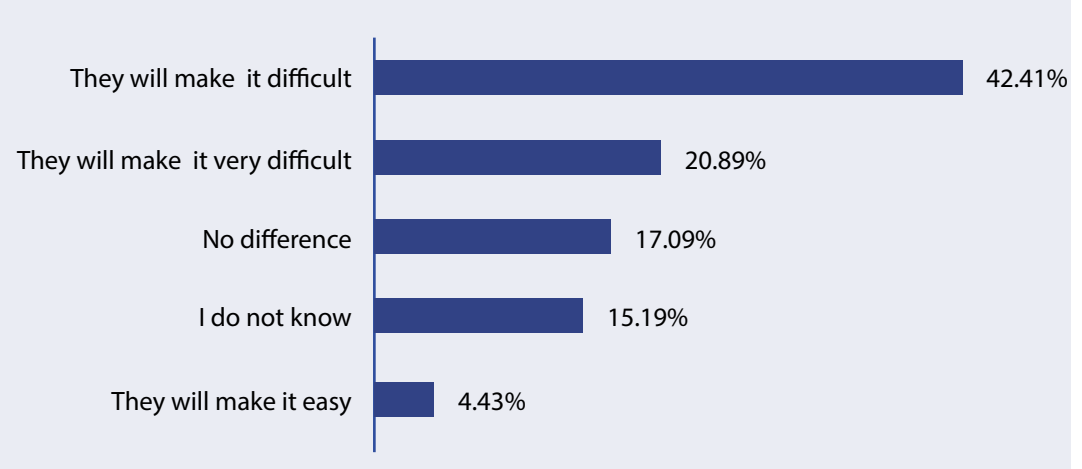


Figure 10: Impact of data minimization regulations on ability of other AI developers to train AI systems in future

Respondents who thought there would be no difference in the impact of data minimization regulations on the ability of other AI developers to train AI models in the future if the regulations remain unchanged accounted for 17.09% of the respondents while 15.19% did not know what the impact would be if the regulations remained unchanged. The remaining 4.43% said that if the regulations remained unchanged, it would make it easy for other AI developers to train AI models.

Respondents further outlined that these regulations would make getting accurate data difficult as it needs more monetary resources that are not adequate to them. They also indicated that without accurate information for the models, AI development would be very challenging.

Data Non-Repurposing

This research further sought to determine whether respondents were knowledgeable on the principle of data non-repurposing. The study defined this principle as not reprocessing already-collected data for secondary use unless one gets consent once again.

Awareness of regulations on data non-repurposing

Figure 11 shows a representation of the respondents' awareness of regulations on data non-repurposing.

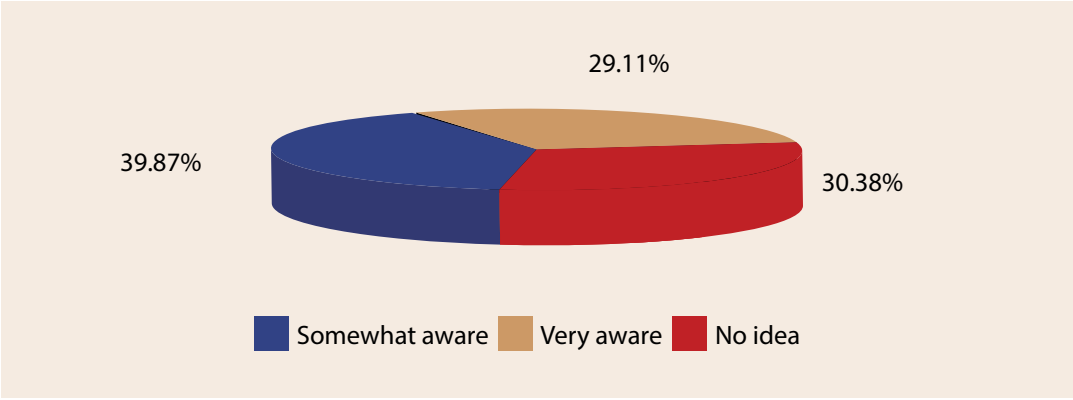


Figure 11: Awareness of regulations on data non-repurposing

Respondents who were very aware of the regulations of data non-repurposing in their countries accounted for 29.11% while those who were somewhat aware contributed to 39.87% of the respondents. The remaining 30.38% of the respondents had no idea on the regulations on data non-repurposing in their fields of AI in their countries.

Of those who were aware of the regulations, most of them (28) said they have made it difficult for them to train AI models, followed by those who did not see a difference (14) in the effect of these regulations on their ability to train AI models. The remaining respondents indicated that the regulations made it easy for them to train AI models (3) and others did not know of the impact of these regulations (1) as shown in table 11.

Table 11: Awareness of data non-repurposing regulations and their impact on ability to train AI models

How aware are you of the existence of such regulations?					Total
How do such regulations affect your ability to train AI models?		No idea	Somewhat aware	Very aware	
I do not know	Count	15	0	1	16
	% within	31.3%	0.0%	2.2%	10.2%
They have made it difficult	Count	23	39	28	90
	% within	47.9%	61.9%	60.9%	57.3%
No difference	Count	10	22	14	46
	% within	20.8%	34.9%	30.4%	29.3%
They have made it easy	Count	0	2	3	5
	% within	0.0%	3.2%	6.5%	3.2%
Total	Count	48	63	46	157
	% within	100.0%	100.0%	100.0%	100.0%

A majority of respondents who had no idea of data non-repurposing regulations (23) said they made it difficult to train models. These were followed by those who did not know whether the regulations had any impact on their training AI models (15). Those who saw no difference of the impact of data non-repurposing regulations on their ability to train AI models yet had no idea of these regulations came in third with 10 respondents. No respondent in this category found it easy to train models.

Some of the respondents were willing to learn about these regulations as they were not aware of them. For those who indicated that they were somewhat aware of the regulations, they said they knew about the Data Protection Act only. In addition to being very aware of the regulations, some respondents showed that they have written research papers on the same.

Impact of data non-repurposing regulations on ability to train AI models

Most respondents (57.59%) feel that regulations on data non-repurposing have made it difficult to train AI models as shown in figure 12.

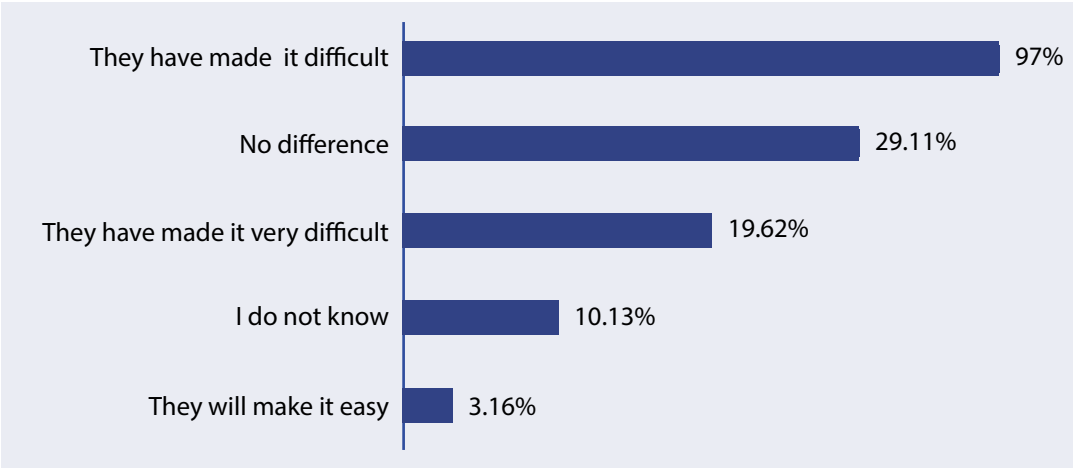


Figure 12: Impact of data non-repurposing regulations on ability to train AI systems

Those who feel that these regulations have made it very difficult account for 19.62% of the respondents while those who feel that it is difficult account for 37.97%. Some respondents (29.11%) feel that there is no difference on the impact of these regulations on their ability to train AI models while 10.13% of the respondents do not know the impact these regulations have on their ability to train AI models. The remaining 3.16% of the respondents said that the regulations on data non-repurposing have made it easy for them to train AI models in their countries of work.

As most respondents indicated that these regulations made it difficult for AI developers to train AI models, respondents further cited that it is usually harder to seek consent for data that has already been collected, despite the consent being for another new use. This is further aggravated by the fact that there are no procedures to record how the data will be processed by the AI develops

hence the temptation to use the already collected data for another purpose. Other respondents said that data that has already been collected might be needed for retraining a model hence seeking multiple consents after makes expensive and very difficult to train AI models. This therefore implies that data that has already been collected cannot be used and whenever historical trends are needed by a model, it is inaccessible and difficult to process the data. On the contrary, some respondents who said the regulations made it easier said it was because every person needs to know how their information is processed and for what purpose and once they understand, seeking consent becomes easier.

The study noted that respondents in the machine learning, data mining and data analysis areas said they are currently affected by the data non-repurposing regulations. Table 12 shows that 60.7% (34) of those in data mining and data analysis and 53.3% (32) of respondents in machine learning said the regulations have made it difficult to train AI models in their areas. The trend was the same for other areas where more than 605 of respondents in their respective categories agreed that the regulations have made it difficult for the m to train AI models. Another significant percentage of respondents indicated that they saw no difference in the impact of data non-repurposing regulations on their ability to train models. The least number of respondents indicated that the regulations on data non-repurposing made it easy for them to train AI models in their areas of focus.

Going by their , most respondents (45) who said that data non-repurposing regulations made it difficult for them to train AI models came from the Fintech field as shown in Table 12. These were followed by those in Education, research, and others with 22 respondents and those from the health field came third (14). Energy, sustainability 7 environmental had 5 respondents, agriculture, 3 and natural language and processing, 2.

Table 12: Effect of data non-repurposing regulations on the training of AI models per field of work

		How do such regulations affect your ability to train AI models?				Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Agriculture	Count	4	3	4	0	11
	% within	36.4%	27.3%	36.4%	0.0%	100.0%
Natural Language Processing & Autonomous Vehicles	Count	1	2	2	0	5
	% within	20.0%	40.0%	40.0%	0.0%	100.0%
Fintech	Count	6	45	25	3	79
	% within	7.6%	57.0%	31.6%	3.8%	100.0%
Health	Count	2	14	7	1	24
	% within	8.3%	58.3%	29.2%	4.2%	100.0%
Energy, sustainability & environmental issues	Count	0	5	1	0	6
	% within	0.0%	83.3%	16.7%	0.0%	100.0%
Education, Research & other	Count	3	22	7	1	33
	% within	9.1%	66.7%	21.2%	3.0%	100.0%
Total	Count	16	91	46	5	158
	% within	10.1%	57.6%	29.1%	3.2%	100.0%

Similarly, most respondents (25) who did not see a difference in the impact of data non-repurposing regulations on their ability to train AI models were from the Fintech field followed by those in health (7) and then education, research, and others (7). This trend was the same for those who indicated that the regulations made it easy for them to train their AI models with a majority (3) being in the Fintech field, followed by health (1) and thirdly, education, research, and others (1).

Given the respondents' area of focus in AI, a majority of (34) respondents who indicated that data non-repurposing regulations made it difficult for them to train their AI models were from the Data Mining and Data Analysis area followed by those in machine learning (32) and the third category were from Natural Language and Processing (9) as shown in table 13.

Table 13: Effect of data non-repurposing regulations on the training of AI models per respondents' area of focus

	How do such regulations affect your ability to train AI models?					Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Robotics	Count	0	3	1	1	5
	% within	0.0%	60.0%	20.0%	20.0%	100.0%
Machine learning	Count	8	32	19	1	60
	% within	13.3%	53.3%	31.7%	1.7%	100.0%
Natural Language Processing	Count	1	9	3	0	13
	% within	7.7%	69.2%	23.1%	0.0%	100.0%
Knowledge Based Systems & Expert Systems	Count	0	3	1	1	5
	% within	0.0%	60.0%	20.0%	20.0%	100.0%
Image analysis and vision	Count	3	8	1	0	12
	% within	25.0%	66.7%	8.3%	0.0%	100.0%
Data Mining and Data Analysis	Count	3	34	17	2	56
	% within	5.4%	60.7%	30.4%	3.6%	100.0%
Others	Count	1	2	4	0	7
	% within	14.3%	28.6%	57.1%	0.0%	100.0%
Total	Count	16	91	46	5	158
	% within	10.1%	57.6%	29.1%	3.2%	100.0%

In the Data mining and Data analysis area of focus, 1.27% of respondents said the regulations made it easy for them to train AI models. For the Robotics, Machine learning, and Knowledge Based Systems & Expert Systems areas of focus, 0.63% of respondents in those respective areas indicated that the regulations made it easy for them to train models.

Respondents (50) who got their data from secondary sources indicated that data non-repurposing

regulations have made it difficult for them to train AI models. These were followed by those whose source was primary data (36) and finally dummies (3) and other sources (2) as shown on table 14. Out of the 3.2% (5) who said that data non-repurposing regulations made it easy for them to train AI models, 3 indicated that they got their data from primary sources and 2 from secondary data.

Table 14: Impact of data non-repurposing regulations on ability to train AI models by respondents' sources of data

	How do such regulations affect your ability to train AI models?					Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Primary data	Count	5	36	20	3	64
	% within	7.8%	56.3%	31.3%	4.7%	100.0%
Secondary data	Count	9	50	23	2	84
	% within	10.7%	59.5%	27.4%	2.4%	100.0%
Dummies	Count	0	3	1	0	4
	% within	0.0%	75.0%	25.0%	0.0%	100.0%
Others	Count	2	2	2	0	6
	% within	33.3%	33.3%	33.3%	0.0%	100.0%
Total	Count	16	91	46	5	158
	% within	10.1%	57.6%	29.1%	3.2%	100.0%

For those who indicated that they saw no difference in the effect of the data non-repurposing regulations on their ability to train AI models, a majority (23) said their source of data was secondary data, followed by those whose source was primary (20). Those who used dummies and other sources of data were 1 and 2 respectively.

In addition to sources of data, the study sought to establish how data non-repurposing regulations affected the respondents by their perspective of criticality of data in their field of AI. Table 15 outlines the effect of these regulations. Most respondents (83) who thought data was highly critical in their field of AI said that the regulations made it difficult for them to train AI models, they were followed by those who said data was not critical in their field but thought that the regulations made it difficult for them to train AI models (8).

Most respondents (35) who saw no difference on the impact of data non-repurposing regulations on their ability to train AI models said data was very critical in their area of AI while the least number (11) thought that data was not critical.

Table 15: Impact of data non-repurposing regulations on ability to train AI models by criticality of data

	How do such regulations affect your ability to train AI models?					Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Not at all	Count	1	2	1	0	4
	% within	25.0%	50.0%	25.0%	0.0%	100.0%
Very little	Count	2	6	10	1	19
	% within	10.5%	31.6%	52.6%	5.3%	100.0%
Quite a lot	Count	2	23	9	3	37
	% within	5.4%	62.2%	24.3%	8.1%	100.0%
Very much	Count	11	60	26	1	98
	% within	11.2%	61.2%	26.5%	1.0%	100.0%
Total	Count	16	91	46	5	158
	% within	10.1%	57.6%	29.1%	3.2%	100.0%

For those who indicated that the regulations made it easy for them to train AI models, a majority (4) indicated that data was very critical in their area while the rest (1) said that data was not critical.

Impact of data non-repurposing regulations on ability to train AI models in future if unchanged

On what respondents thought about the effect of regulations on data non-repurposing, if they remained as they are, on their ability to train AI models in the future, most of them (70.26%) said it would make it very difficult as shown on figure 13.

Figure 13: Impact of data non-repurposing regulations on ability to train AI models in future

Some respondents accounting for 17.09% indicated that there would be no difference in the impact of regulations on data non-repurposing on their ability to train AI models in the future if the regulations remained as they are. A further 8.86% did not know whether there would be any impact on their ability and the remaining 3.16% thought that if the regulations remained unchanged, their ability to train AI models would be easy.

The study noted that respondents felt that if regulations remain unchanged, it would be difficult to have data for training models. The respondents added that in cases where already collected data forms a basis for another model and given the scarcity of data, these regulations would limit developers in training AI models.

Moreover, the study wanted to establish how respondents in their various fields of work thought data non-repurposing regulations would affect their ability to train AI models in future if the regulations remained unchanged. This is outlined in table 16. Slightly over half of these respondents (58) were from the Fintech field, followed by Education, research & others (24) and then health (16). The agriculture, natural language and processing and Energy, sustainability & environmental fields had the least respondents with 6, 5 and 2 respondents respectively.

Table 16: Impact of data non-repurposing regulations on ability to train AI models in future if unchanged by field of work

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Agriculture	Count	3	6	2	0	11
	% within	27.3%	54.5%	18.2%	0.0%	100.0%
Natural Language Processing & Autonomous Vehicles	Count	1	2	1	1	5
	% within	20.0%	40.0%	20.0%	20.0%	100.0%
Fintech	Count	8	58	11	2	79
	% within	10.1%	73.4%	13.9%	2.5%	100.0%
Health	Count	1	16	4	3	24
	% within	4.2%	66.7%	16.7%	12.5%	100.0%
Energy, sustainability & environmental issues	Count	0	5	1	0	6
	% within	0.0%	83.3%	16.7%	0.0%	100.0%
Education, Research & other	Count	1	24	8	0	33
	% within	3.0%	72.7%	24.2%	0.0%	100.0%
Total	Count	14	111	27	6	158
	% within	8.9%	70.3%	17.1%	3.8%	100.0%

Those respondents who indicated that there would be no impact on their ability to train models if these regulations remained unchanged dropped to 27 with the majority being from the Fintech field followed by Education, research and other. Those who said that it will be easy for them to train AI models despite the data non-repurposing regulations remaining the same in the future stood at 6.

Given the respondents' area of focus in AI, the study established that most respondents (43) in data mining and data analysis said that if the current data non-repurposing regulations remain unchanged, it will be difficult to train AI models in the future. These we followed by respondents in machine learning (41) and image analysis and vision and Natural language processing with 9 respondents each indicating the same view of difficulty as shown on table 17.

Table 17: Impact of data non-repurposing regulations on ability to train AI models in future if unchanged per area of focus in AI

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Robotics	Count	0	3	1	1	5
	% within	0.0%	60.0%	20.0%	20.0%	100.0%
Machine learning	Count	9	41	10	0	60
	% within	15.0%	68.3%	16.7%	0.0%	100.0%
Natural Language Processing	Count	1	9	2	1	13
	% within	7.7%	69.2%	15.4%	7.7%	100.0%
Knowledge Based Systems & Expert Systems	Count	0	4	0	1	5
	% within	0.0%	80.0%	0.0%	20.0%	100.0%
Image analysis and vision	Count	2	7	2	1	12
	% within	16.7%	58.3%	16.7%	8.3%	100.0%
Data Mining and Data Analysis	Count	2	43	9	2	56
	% within	3.6%	76.8%	16.1%	3.6%	100.0%
Others	Count	0	4	3	0	7
	% within	0.0%	57.1%	42.9%	0.0%	100.0%
Total	Count	14	111	27	6	158
	% within	8.9%	70.3%	17.1%	3.8%	100.0%

Most of those who indicated that it will be easy to train AI models if the regulations remain unchanged in future (2) were from data mining and data analysis, followed by the rest of the categories with 1 respondent each.

Most respondents (61) who got their data from secondary sources indicated that if the data non-repurposing regulations remained unchanged, it would make training of AI models difficult in the future. This was followed by those whose source was primary data (42) and those who used dummies and other sources at 4 respondents each as shown on table 18.

Table 18: Impact of data non-repurposing regulations if they remain unchanged on respondents' ability to train AI models in the future by source of data

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
		I do not know	They have made it difficult	No difference	They have made it easy	
Primary data	Count	5	42	15	2	64
	% within	7.8%	65.6%	23.4%	3.1%	100.0%
Secondary data	Count	8	61	11	4	84
	% within	9.5%	72.6%	13.1%	4.8%	100.0%
Dummies	Count	0	4	0	0	4
	% within	0.0%	100.0%	0.0%	0.0%	100.0%
Others	Count	1	4	1	0	6
	% within	16.7%	66.7%	16.7%	0.0%	100.0%
Total	Count	14	111	27	6	158
	% within	8.9%	70.3%	17.1%	3.8%	100.0%

Most respondents (4) who said they thought it would be easy for them to train AI models in the future if data non-repurposing regulations remained unchanged got their data from secondary sources while the rest (2) got it from primary sources.

The study further sought to establish the perspective of respondents on the impact of data non-repurposing regulations on their ability to train AI models in the future if the regulations remain unchanged. Table 19 shows that most respondents (98) who indicated that data is very critical in their area will find it difficult to train AI models in future if the regulations remain unchanged. The rest (13) who found data to be not critical indicated that it would be difficult to train AI models in future if regulations remained unchanged.

Table 19: Impact of data non-repurposing regulations if they remain unchanged on respondents' ability to train AI models in the future by data criticality

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?						Total
Do you consider access to data to be a critical issue in the field of AI which you are in?		I do not know	They have made it difficult	No difference	They have made it easy	
Not at all	Count	1	3	0	0	4
	% within	25.0%	75.0%	0.0%	0.0%	100.0%
Very little	Count	3	10	6	0	19
	% within	15.8%	52.6%	31.6%	0.0%	100.0%
Quite a lot	Count	0	24	10	3	37
	% within	0.0%	64.9%	27.0%	8.1%	100.0%
Very much	Count	10	74	11	3	98
	% within	10.2%	75.5%	11.2%	3.1%	100.0%
Total	Count	14	111	27	6	158
	% within	8.9%	70.3%	17.1%	3.8%	100.0%

Conversely, all respondents (6) who found data to be very critical in their area of focus said that it would be easy for them to train AI models in future if the regulations remained unchanged.

The study noted that most respondents (29) who said they were aware of data non-repurposing regulations indicated that if they remained unchanged, it will be difficult for them to train AI models in the future as shown on table 20. These were followed by those (11) who said there would be no difference in the impact on training AI models if the regulations remained unchanged in the future. The remaining respondents (6) who were aware of data non-repurposing regulations indicated that they would not know (3) the impact of the regulations or it would make it easy (3) for them to train AI models in the future if the regulations remained unchanged.

Table 20: Awareness of data non-repurposing regulations and their impact on respondents' ability to train AI models in the future if they remain unchanged

If they stay as they are, how do you think such regulations will affect your ability to train AI models in the future?					Total
		No idea	Somewhat aware	Very aware	
I do not know	Count	8	3	3	14
	% within	16.7%	4.8%	6.5%	8.9%
They have made it difficult	Count	37	44	29	110
	% within	77.1%	69.8%	63.0%	70.1%
No difference	Count	3	13	11	27
	% within	6.3%	20.6%	23.9%	17.2%
They have made it easy	Count	0	3	3	6
	% within	0.0%	4.8%	6.5%	3.8%
Total	Count	48	63	46	157
	% within	100.0%	100.0%	100.0%	100.0%

For those who were not aware of the data non-repurposing regulations, a majority (37) indicated that it would make it difficult for them to train AI models in the future if the regulations remined unchanged. These were followed by those who did not know the impact (8) and those who saw no difference (3) on the impact of these regulations if they remained unchanged in future.

Impact of data non-repurposing regulations on ability of other AI developers to train AI models

It was important for the study to establish the perspective of the respondents on the impact of data non-repurposing regulations on the ability of other AI developers to train AI models in their countries and the results are shown in figure 14.

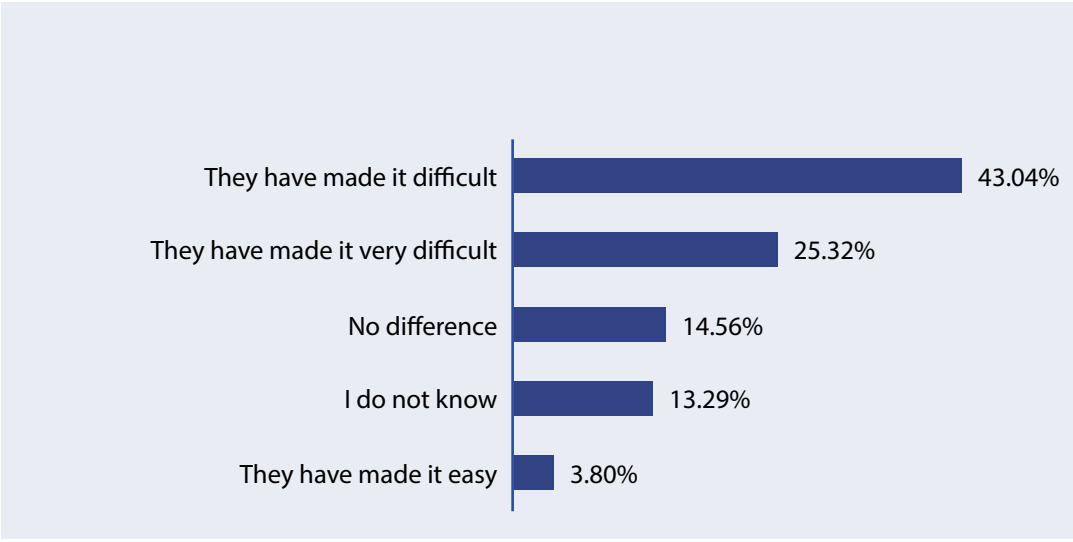


Figure 14: Impact of data non-repurposing regulations on ability of other AI developers to train AI models

Findings showed that most respondents (68.36%) thought that regulations on data non-repurposing have made it difficult for other developers to train AI models in their fields of focus in their countries. For respondents who thought it would be difficult, they indicated that it would be difficult to keep mining new data for models as not all datasets are easily accessible. Others said this would lead to AI systems not being super intelligent hence making it difficult for development.

Moreover, 14.56% of the respondents said they did not expect any difference on the impact of these regulations on the ability of other AI developers to train AI models while 13.29% indicated that they did not know whether the regulations had any impact on the ability of other AI developers' ability to train AI models. The remaining 3.8% of respondents thought that the regulations made it easy for other Ai developers to train AI models.

Impact of data non-repurposing regulations on ability of other AI developers to train AI models in future if unchanged

The study also wanted to establish what respondents thought would be the impact of regulations on data non-repurposing on the impact of the ability of other AI developers to train AI models in the future if the regulations remained unchanged. The results are illustrated in figure 15.

Figure 15: Impact of data non-repurposing regulations on ability of other AI developers to train AI models in future if unchanged

Most respondents (69.62%) felt that if they remain unchanged, regulations on data non-repurposing would make it difficult for other AI developers to train AI models in their fields while 16.46% of the respondents felt that there would be no difference on the impact on the ability of other AI developers training AI models. A further 10.13% of the respondents said they did not know whether there would be any impact on the ability of other AI developers to train AI models in future if the regulations remained unchanged while the remaining 3.8% thought it would be easy for other AI developers to train AI models in future.

Respondents outline those regulations that would inhibit data collection in the future would create a hurdle in their ability to train data. They also that since not all datasets they require are easily accessible, it would be difficult to mine data in future. Some recommendations they gave included finding better ways to collect data legally and pushing for legislation that would enhance open banking and open APIs in the AI field.

This study was made possible by a grant provided by the International Development Research Center (IDRC). We thank the organization for their continued support.

Canada