

Transcript | AI Skills Lab | AI for Analysing Data

WEBVTT

1

00:00:04.440 --> 00:00:12.779

Joel Smith: Hello! Hello, everybody, and welcome to this, the first edition of the AI Skills Lab.

2

00:00:12.860 --> 00:00:29.889

Joel Smith: Covering AI for analyzing data. My name is Joel Smith, and I am an events manager here at Apolitical. Now, the AI Skills Lab is a new monthly virtual event series brought to you as part of our Government AI Campus.

3

00:00:29.890 --> 00:00:45.090

Joel Smith: Bringing you tangible tips and hacks on how to streamline your everyday tasks with AI. Now, it would be great to hear from you all, so please do post in the chat where you're joining us from and what organization you work for.

4

00:00:45.180 --> 00:00:57.690

Joel Smith: Now today, we are going to be covering how to use AI tools to help you clean, interpret, and visualize data faster, and make better decisions with less manual effort.

5

00:00:57.940 --> 00:01:12.269

Joel Smith: Now, for those of you who may be joining us for the first time, welcome. Apolitical is the world's largest community for government, used by over a quarter of a million public servants and policymakers from more than 170 countries.

6

00:01:12.330 --> 00:01:21.790

Joel Smith: Our mission is to make governments smarter, and we do this by equipping public servants worldwide with the knowledge and skills they need to lead the change.

7

00:01:22.040 --> 00:01:37.509

Joel Smith: Now, I'll just jump into the chat and see who is introducing themselves. So, we have, we have people dialing in from all over. We've got Canada, we've got, Ontario, also Canada.

8

00:01:37.510 --> 00:01:50.289

Joel Smith: Joining from... Susie, joining from the Immigration Advice Authority UK. Hello, welcome. Sarah, joining from DFE in London. We have...

9

00:01:50.350 --> 00:02:02.250

Joel Smith: Tasneem joining from the Department of Health, also London, lots of London and Canada, which is quite usual on our events. Jack, I see you dialing in from Dublin, Ireland, welcome!

10

00:02:02.250 --> 00:02:15.579

Joel Smith: Lovely to see all of you here, thank you, and do keep getting your messages in the chat, we love to see where you're all kind of dialing in from, and yeah, nice for you to kind of join us. We want this to be nice and interactive.

11

00:02:15.800 --> 00:02:32.220

Joel Smith: Today's session is going to be a very compact 30 minutes, tightly focused on building your AI skills. We know that as public servants, you are likely short on time, so we want these events to suit your learning needs and work with your busy schedules.

12

00:02:32.420 --> 00:02:46.420

Joel Smith: Now, we're going to get to the presentation very shortly, but just before we do, we are thrilled to extend an invitation to join nearly 12,000 of your peers in public service from across the globe in our AI in Government community.

13

00:02:46.420 --> 00:02:53.999

Joel Smith: Now, this is an online space to share conversations, insights, ideas, and resources on the use of AI in government.

14

00:02:54.000 --> 00:03:15.969

Joel Smith: through connection and conversation with peers across the globe. And this space is open to all, whether you consider yourself an AI expert or an AI novice. Now, the link for that is in the resources tab, which should be at the bottom of your screen somewhere, and my colleague will also post a link in the chat, so we would love to see your voices in there after this event.

15

00:03:16.170 --> 00:03:27.500

Joel Smith: I have a very quick technical note to get out the way before we get into the session. So, this event is recorded for on-demand viewing on the apolitical website, so please don't record it yourself.

16

00:03:27.530 --> 00:03:39.600

Joel Smith: Contact details for any privacy concerns are in the apolitical homepage footer, and closed captions are available, and they can be toggled on or off in your screen settings.

17

00:03:39.920 --> 00:03:41.010

Joel Smith: Okay.

18

00:03:41.010 --> 00:04:00.309

Joel Smith: Now, on to the main event. It is my pleasure to welcome my colleague, Jesse Samasuo, who is our speaker for today. Jesse is our Senior Policy and Insights Manager here at Apolitical. His work focuses on researching and understanding the capabilities governments need.

19

00:04:00.310 --> 00:04:04.680

Joel Smith: To build to be effective in a world shaped by digital technologies.

20

00:04:04.680 --> 00:04:23.889

Joel Smith: He led the development of the UK government's Digital Excellence Programme for senior civil servants, the 2024 State of Learning in Government report involving 2,000 public servants, and most recently, insights on government AI readiness based on data from 5,000 public servants globally.

21

00:04:24.180 --> 00:04:43.129

Joel Smith: So, Jesse, it is an absolute pleasure to have you online with us. I'm going to hand over to you for the next 15 minutes or so, and then we're going to have a few minutes reserved for Q&A. So, attendees, while Jesse is speaking, feel free to submit your questions in the Q&A tab at the bottom of your screen, but Jesse, the floor is yours.

22

00:04:43.450 --> 00:04:52.709

Jesse Samasuwo: Thanks, Joel. Hi, everyone. My name is Jesse, as Joel said, I'm the Senior Manager for Policy and Insights here at Apolitical, and really looking forward to spending the next 15 minutes talking to you about

23

00:04:52.710 --> 00:05:04.650

Jesse Samasuwo: Some trends on how we're seeing governments make use of AI, specifically generative AI for analyzing data, and how you can adopt some of these methods, hopefully in your own work. So, let's start by sharing screen.

24

00:05:04.880 --> 00:05:10.470

Jesse Samasuwo: And we will get going. Let me know if you can see that, Joel.

25

00:05:10.590 --> 00:05:12.700

Jesse Samasuwo: As planned? All good.

26

00:05:12.700 --> 00:05:16.920

Joel Smith: It's just loading up, but should be, should be on shortly.

27

00:05:23.070 --> 00:05:25.429

Jesse Samasuwo: Coming through now, I think I can see a thumbs up.

28

00:05:27.410 --> 00:05:33.130

Joel Smith: We're getting some thumbs up. I can't see it on my end, but the audience are giving a few thumbs up. There we are!

29

00:05:33.130 --> 00:05:35.160

Jesse Samasuwo: Okay, perfect.

30

00:05:35.330 --> 00:05:40.719

Jesse Samasuwo: So yeah, so we'll be talking today for the next 15 minutes or so about how you can use AI for analyzing data in your work.

31

00:05:41.480 --> 00:05:58.269

Jesse Samasuwo: Worth stating up front, all of the available evidence and sort of research that's out there suggests that this isn't actually a commonly done thing, so people in government and public servants don't tend to be using AI for data analysis. There are lots of really good reasons for that, some of which we'll get into throughout our conversation.

32

00:05:58.270 --> 00:06:07.200

Jesse Samasuwo: But if you're feeling as though you're not quite using AI in this way, don't feel as though you're missing out on everything... missing out on something, because the available evidence suggests that not many people do.

33

00:06:07.200 --> 00:06:19.150

Jesse Samasuwo: Only around 4% when we asked public servants around 12 months ago, and more recent surveys have shown that it's around that same figure. So, worth stating that this is actually you at the cutting edge of trying to think about how to use AI in your work.

34

00:06:19.380 --> 00:06:24.570

Jesse Samasuwo: But that's not to say that there aren't opportunities for using AI for data analysis.

35

00:06:24.690 --> 00:06:36.269

Jesse Samasuwo: We are definitely seeing a growing trend of governments making greater use of AI to process, manage, and analyze information, and we'll be talking about how you can make some use of some of these methods in your own work.

36

00:06:36.370 --> 00:06:45.810

Jesse Samasuwo: So by the end of this session, hopefully you'll have some practical ideas, to help you to start thinking about how you can incorporate some of these methods in your day-to-day work.

37

00:06:46.520 --> 00:07:04.899

Jesse Samasuwo: Another thing to... another thing worth saying up front is that, let's see if we can... there we go. For data analysis, you want to make sure that the things that... the work that you're doing is systematic and reproducible. That's really important because it's, helps you build credibility and also makes sure that your work is externally, can be validated by your colleagues.

38

00:07:04.900 --> 00:07:21.720

Jesse Samasuwo: So when we're working through some of these examples, you'll see that we're talking about using AI really as a tool to support the production of analytical outputs, rather than using it as a... to wholly automate outputs on its own. And I think that's, like, an important distinction that you'll see kind of come through in a lot of our notes today.

39

00:07:22.420 --> 00:07:30.970

Jesse Samasuwo: So, what does that mean in practice? When we're thinking about systematic approaches rather than automated analysis, you can see on the screen, hopefully on the left, two examples. So.

40

00:07:31.130 --> 00:07:49.019

Jesse Samasuwo: I've downloaded an open dataset that we have at Apolitical here. It's a survey from 2024, looking at around how 600 public servants are using AI in government. Now, I've just uploaded that to an AI tool and asked it to provide a breakdown of the age group of respondents, and you can see on the left that it's given me that data.

41

00:07:49.110 --> 00:08:07.150

Jesse Samasuwo: On the right-hand side, instead, I've asked to talk me through how I could create something called a pivot table to analyze that same data, and it's... you can see the beginning of that chat, where it's starting to give me the instructions, and then I've implemented those in Google Sheets to kind of generate that pivot table on the right.

42

00:08:07.150 --> 00:08:19.170

Jesse Samasuwo: So you can see that it's exactly the same output, but the methods are different. One is kind of what I would describe as a wholly automated analysis, and the other is more like a systematic approach, where AI is kind of talking you through a process.

43

00:08:19.170 --> 00:08:38.570

Jesse Samasuwo: So what we'll be focusing on today is more about the second half of that take... the second half of that picture, rather than the left-hand picture, and I would encourage you to think about using AI as a tool to help you do the kind of work on the right, rather than trying to use it to completely automate analysis, because I think that builds trust and kind of reliability in the way you're working with AI.

44

00:08:39.100 --> 00:08:54.499

Jesse Samasuwo: So, this is a visualization that you'll see a couple of times in the next 10 minutes or so, and it's from an upcoming piece of research we're about to publish. We looked at just over 300 different generative AI projects in governments from around the world, from South Korea to Singapore to the United States to Finland.

45

00:08:54.500 --> 00:09:03.570

Jesse Samasuwo: And we looked at how they are using AI in their operations. A subset of those projects, around 50 or so, are specifically focused on data operations.

46

00:09:03.570 --> 00:09:18.039

Jesse Samasuwo: And what you can see in front of you is a kind of data cycle, the life cycle that a piece of information goes through from being generated all the way through to being analyzed, visualized, interpreted. And you can see how governments are making use of AI across that data life cycle.

47

00:09:18.040 --> 00:09:27.610

Jesse Samasuwo: And we'll be focusing, really, on these three biggest ones, so processing, management, and analysis. And that's what we... those will be some of the kind of ways we talk about using AI in your work.

48

00:09:27.710 --> 00:09:31.200

Jesse Samasuwo: So what do we mean by that? So when we talk about AI as a data processor.

49

00:09:31.480 --> 00:09:39.460

Jesse Samasuwo: We mean helping... we mean using AI to help you structure and organize datasets more efficiently, so you can use them, so you can analyze them more effectively.

50

00:09:39.460 --> 00:09:49.720

Jesse Samasuwo: When we're talking about AI as a data manager, we mean by helping you... we mean helping... using AI to help you gain access to new sources of data and find and recruiting them more easily, so you can use it in your work.

51

00:09:49.720 --> 00:10:00.089

Jesse Samasuwo: And then using AI as a data analyzer, this is kind of what we were talking about earlier around sort of, like, using AI to help automate some of the analysis of data in specific ways, which we will describe later on.

52

00:10:00.090 --> 00:10:09.299

Jesse Samasuwo: This is a short event, so we'll try to get through as much as we can, but if there's anything that we miss out, I'm sure we'll be sharing my slides, and we can sort of pick up the conversation in the community.

53

00:10:09.530 --> 00:10:18.519

Jesse Samasuwo: So let's start with processing data. So how can you use a large language model to help you organize a data set for easier analysis?

54

00:10:18.840 --> 00:10:37.490

Jesse Samasuwo: And this is something that we're definitely seeing governments do a lot of, specifically in the area around... specifically in the area of public consultations, which I'm sure lots of you work with in your day-to-day work. These are huge text-heavy datasets, which governments need to be able to use more efficiently, and we're seeing large language models helping them accelerate that process.

55

00:10:37.490 --> 00:10:40.130

Jesse Samasuwo: And you can adopt some of these methods in your own work.

56

00:10:40.650 --> 00:10:58.360

Jesse Samasuwo: What we're actually going to go through for the next couple of minutes is a kind of, like, a worked example of something that we did here at AP Political to generate that piece of research I was just showing you. So what you can see here is a screenshot of the 2024 Federal AI Use Case Inventory, which is published every year by the United States government at the federal level.

57

00:10:58.440 --> 00:11:17.049

Jesse Samasuwo: This is a huge dataset of around 2,000 different AI projects from across the federal government, and what we have here is, basically a kind of classic processing task. So what we want to do is, we want to do two things. We want to first extract all of the projects from this dataset, which only focus on generative AI,

58

00:11:17.050 --> 00:11:29.369

Jesse Samasuwo: And then we want to start tagging those projects so that we can more easily categorize and analyze them for use in our kind of data analysis. So those are kind of two ways in which we want to make sense of this dataset.

59

00:11:29.900 --> 00:11:49.829

Jesse Samasuwo: So, just to illustrate again why we're using... why LLMs are really useful in this context, because of their foundational training, large language models have the capacity to make... to interpret in complex... more... complex and unpredictable linguistic data sets. Machine learning approaches have been used in this way for a long time, but these can often be quite limited when it comes to unpredictable or

60

00:11:49.910 --> 00:12:02.690

Jesse Samasuwo: less well-formulated data, but large language models, because of their extensive training, can kind of adapt to new and unseen information much more effectively, and that's what you'll be seeing in the next few slides. So...

61

00:12:02.710 --> 00:12:12.750

Jesse Samasuwo: as we said, we've got this dataset, and we want to extract two things. We want to extract the generative AI projects, and then we want to start tagging them so that we can understand and analyze them more easily. So what does that look like?

62

00:12:12.750 --> 00:12:32.020

Jesse Samasuwo: Well, really, it's kind of a two-step process, one which relies on large language abilities, large language abilities, and one which doesn't. So you can see on the left here, what we want to do in that first step is basically just a keyword search. We want to systematically extract all of the projects which use some of these key terms, like generative AI, LLM,

63

00:12:32.160 --> 00:12:51.410

Jesse Samasuwo: RAG, Copilot, all of these kind of indicative words that suggest this might be a generative AI project, and extract them and put them into a new sort of Google Sheet or Excel sheet that helps us kind of just analyze that data more effectively. Now, this is just a classic keyword search. You could do it manually if you wanted in Excel or whatever tool you're using.

64

00:12:51.410 --> 00:12:53.010

Jesse Samasuwo: But it's just a way to show that

65

00:12:53.020 --> 00:13:11.909

Jesse Samasuwo: it's a kind of first step which doesn't require an LLM ability. On the right-hand side, what you can see is a screenshot of a sort of Python script, part of which includes an AI prompt, and you can read that maybe a little bit. If you zoom in a little bit, you can see that it says.

66

00:13:11.910 --> 00:13:28.020

Jesse Samasuwo: You're an AI research assistant analyzing a database of government generative AI projects, and it asks the LLM to tag all of those projects. So to read the project description that you see here in these last two columns, and then add a new column where there's a short, concise tag to describe that project.

67

00:13:28.940 --> 00:13:29.880

Jesse Samasuwo: Now...

68

00:13:30.280 --> 00:13:38.210

Jesse Samasuwo: I could have done that manually, we could have done that manually, and I'm sure you've all had the experience of trying to upload, sort of, like, a huge piece of information into whatever tool you're using.

69

00:13:38.270 --> 00:13:47.150

Jesse Samasuwo: And getting, the LLM to... to do whatever you want. In this case, tag the projects.

70

00:13:47.180 --> 00:13:59.910

Jesse Samasuwo: that is an approach that can work, but there are obviously some important limitations there. You know, there are context... what are called context windows, which means that there's only so much information that the LLM can process, until it starts to, sort of.

71

00:13:59.910 --> 00:14:05.639

Jesse Samasuwo: Get a bit disoriented and start to, you know, increase the risk of hallucination and sort of, like, less reliable outputs.

72

00:14:05.640 --> 00:14:29.939

Jesse Samasuwo: So, I could have manually done this process either by uploading, copying and pasting lots of that text at once, which probably would have led to quite a short... to less high-quality outputs, or I could have manually copied and pasted each row and sort of, like, said, okay, here's the next project, tag this. Now, again, that's doable, but obviously not a very efficient process. What we see on the right-hand side is what I would describe as, like, a programmed approach to prompting.

73

00:14:30.210 --> 00:14:33.780

Jesse Samasuwo: Where we are writing a piece of script, which is what we see here.

74

00:14:33.920 --> 00:14:40.539

Jesse Samasuwo: Which instructs the LLM to kind of how... what... instructs the LLM to... on how to behave.

75

00:14:40.540 --> 00:14:54.649

Jesse Samasuwo: And in this instance, we are asking it to read each row one at a time, read that project description, and generate a tag. And you can see it's moving through here around 283 different projects, and reading each of them, and processing each of them, and generating that tag.

76

00:14:54.780 --> 00:15:01.019

Jesse Samasuwo: So, the time was spent writing the script and generating that script, but then the program, once that script has been written.

77

00:15:01.220 --> 00:15:16.019

Jesse Samasuwo: you kind of fire it off, and then the LLM is working through that process for about 8 to 10 minutes, systematically, in the background, basically automating that manual process, of me kind of trying to, you know, individually upload each and every project.

78

00:15:16.670 --> 00:15:38.860

Jesse Samasuwo: So, what does that process look like at the end? We started off, obviously, as I said, with that huge dataset of over 2,000 projects, and we only wanted to extract the generative AI ones, and we ended up with, at the end of this process, a much more refined dataset, with only generative AI projects, and you can see in that yellow column what's described as the purpose tag there, and that tag has been generated by AI.

79

00:15:38.860 --> 00:15:40.190

Jesse Samasuwo: Now, obviously.

80

00:15:40.320 --> 00:15:53.610

Jesse Samasuwo: Once you're working with generative AI or large language models for any kind of analysis process or any kind of work, you obviously introduce the risk of, you know, the well-known risks of hallucination, of oversimplification, and just general unreliability.

81

00:15:53.610 --> 00:16:07.050

Jesse Samasuwo: I think someone once described AIs as sort of incredibly powerful, but also unpredictably unreliable. So whenever you're doing this kind of analysis work, you're trying to minimize the potential risk of that unpredictability, or that unreliability.

82

00:16:07.050 --> 00:16:17.580

Jesse Samasuwo: So in this instance, because it was a relatively small database of around 300 projects, I kind of went through each of these and read through them individually, just to make sure that the tag was accurate.

83

00:16:17.720 --> 00:16:34.980

Jesse Samasuwo: That obviously took time, but it was obviously... it's a much faster process than me trying to read each of the 300 projects and write my own tag, right? So, it's still taking time, I'm still investing time to ensure there's some kind of quality control, but it's a much faster approach than trying to generate that process completely manually on my own.

84

00:16:34.980 --> 00:16:51.469

Jesse Samasuwo: So that's an example of how you can kind of use AI to process a dataset, and now we have these tags. It makes the kind of analytical work much easier. So we're not working with these long descriptions, we're just working with these short tags, which is a much more concise way to work with a dataset.

85

00:16:51.940 --> 00:16:55.690

Jesse Samasuwo: So that's the first example of thinking about how to use AI for

86

00:16:55.870 --> 00:17:10.209

Jesse Samasuwo: Data processing. So let's move now to management. And again, this is another area where we see governments make more and more use of AI, specifically around making it easier for folks in government to find and, you know, retrieve information really quickly.

87

00:17:10.460 --> 00:17:20.890

Jesse Samasuwo: I'm sure all of you know far better than me that government is, like, you know, huge amounts of text-based data that's... that live in government, and it's often been quite hard to make that more accessible.

88

00:17:20.990 --> 00:17:34.819

Jesse Samasuwo: Large language models can help make that information more easily retrievable, and we're seeing governments making use of new tools to make it easier to access data. There are two ways in which this is happening, one which is,

89

00:17:35.620 --> 00:17:49.509

Jesse Samasuwo: creating tools which enable public servants, perhaps who don't have a technical background, to generate, to auto-generate, what are called scripts. So these are short snippets of computer programming language that make it easier for you to navigate,

90

00:17:49.510 --> 00:18:01.970

Jesse Samasuwo: a large data set, and sort of pull out the piece of information that you want, which is what we'll show in a second. And then the other example, as I described earlier, is just kind of, like, making it easier to search, basically, using natural language.

91

00:18:01.970 --> 00:18:04.600

Jesse Samasuwo: Through large databases to find relevant information.

92

00:18:05.680 --> 00:18:20.949

Jesse Samasuwo: So, what we're talking about... what we're looking at here is, a new... a method that you can use in your own work that governments are also using, to get access to what I would describe as novel types of data through something called application programming interfaces. I'm sure

93

00:18:20.980 --> 00:18:30.170

Jesse Samasuwo: if you have heard this terminology before, you'll be familiar with it. If not, the simplest way to think about them is a kind of set of rules or documentation that enable

94

00:18:30.210 --> 00:18:34.279

Jesse Samasuwo: computer systems to communicate with each other.

95

00:18:34.280 --> 00:18:51.809

Jesse Samasuwo: And so we see APIs being used more widely in governments to enable seamless information sharing across departments, or even within and outside of government. Again, these are methods that you can use at the team level, or as an individual to make... to get access to more... to new types of data in your work.

96

00:18:51.960 --> 00:19:02.299

Jesse Samasuwo: You can see on the right two examples of, what's called an API, one from Facebook and one from Reddit, and we'll show how government's actually using this in their own work.

97

00:19:03.170 --> 00:19:19.409

Jesse Samasuwo: So, in the United States, the Department of Health and Human Services, two really, really cool use cases of how they've used APIs to make it easier to do the work. So, during COVID, obviously they want to track school closures and potential outbreaks.

98

00:19:19.410 --> 00:19:25.739

Jesse Samasuwo: Obviously, that started off as quite a manual process, calling hundreds, if not thousands, of school districts across the country to be like.

99

00:19:25.740 --> 00:19:35.709

Jesse Samasuwo: Is your school open or closed today? Have you had any, COVID outbreaks? Obviously an extremely time, time, sort of, like, intensive and manual process.

100

00:19:36.000 --> 00:19:49.769

Jesse Samasuwo: An alternative approach to that is, all of these schools have public Facebook pages, right, where they're publishing information about, for parents, using the API, they were able to access 40,000 public profiles to track

101

00:19:49.770 --> 00:20:01.429

Jesse Samasuwo: school closures, and sort of automate that process of analysis. Again, using some of the methods that we were talking about earlier, using large language models to process that text data and interpret it. Similarly.

102

00:20:01.430 --> 00:20:15.959

Jesse Samasuwo: you know, to help... to understand people's help-seeking behavior for sexual health issues. They used Reddit sub... they used subreddits to understand, how people, you know, make the decision to go for help in that circumstance. So the basic concept here is that there's...

103

00:20:15.960 --> 00:20:30.429

Jesse Samasuwo: information, especially in the internet area, where there's lots of user-generated content online, you can start to get access to some of this information through APIs to help you improve the sort of richness of information that you're using to make decisions and inform some of the work that you're doing.

104

00:20:30.780 --> 00:20:49.880

Jesse Samasuwo: I'll talk through a really, really basic example that, you know, I used in my own work recently. So this is an example of a short Python scripts, again, generated by AI, to help me access a public database of research and track interest in a topic over time. So we were writing a short piece on trying to understand how research interest in public procurement has grown over time.

105

00:20:49.970 --> 00:21:06.340

Jesse Samasuwo: There's a public database called OpenAlex, which has its own API. Again, this is a short script that pulls... that does a quick search of that database, and then pulls all the information about publications related to public procurement over the last 25 years. And really, really quickly, you've got

106

00:21:06.470 --> 00:21:12.829

Jesse Samasuwo: An interesting snapshot, or a way of trying to demonstrate, the interest in this topic over time.

107

00:21:13.150 --> 00:21:22.900

Jesse Samasuwo: So it's like a really... again, this is not using AI to automate the analysis, it's more of, like, a systematic approach where I am... well, we're pulling information from a reliable source.

108

00:21:23.210 --> 00:21:42.249

Jesse Samasuwo: in a systematic way, and then using that to generate a visualization, rather than just asking AI, for example, you know, how much interest has there been in public procurement over time? We're still... we're actually using a reliable source, but we're just using AI to help us get access to it. Sort of reinforcing that... reinforcing that systematic approach again.

109

00:21:42.720 --> 00:21:46.079

Joel Smith: Sorry, Jesse, just to warn you, you've got about one more minute.

110

00:21:47.090 --> 00:21:52.189

Jesse Samasuwo: Perfect. Let's quickly run through the last one, which is using AI for data analysis.

111

00:21:52.360 --> 00:21:59.600

Jesse Samasuwo: Again, same methods and approaches here, using large language models to interpret and make sense of text data.

112

00:21:59.870 --> 00:22:16.819

Jesse Samasuwo: The biggest area we've seen this in governments is, around the use of AI again in consultations. So, for example, trying to evaluate, you know, sentiment in consultations, which is sort of huge amounts of text data.

113

00:22:16.820 --> 00:22:30.230

Jesse Samasuwo: Again, the risks of AI here, the risks of hallucination, oversimplification, repetition, and overlooking nuance. So I just want to quickly highlight how you can sort of try to overcome some of those approach... some of those limitations in your work.

114

00:22:30.450 --> 00:22:41.259

Jesse Samasuwo: And thinking really carefully about curating data. So this is an approach called... you can either call it chunking or context engineering, but the basic concept here is that you're trying to

115

00:22:41.900 --> 00:22:50.339

Jesse Samasuwo: break down the piece of information that you want a large language model to use into more manageable chunks. So here at Apolitical, if we're ever doing interviews.

116

00:22:50.560 --> 00:22:56.500

Jesse Samasuwo: If we're using AI to process those interviews, obviously with user consent explicitly.

117

00:22:56.540 --> 00:23:16.670

Jesse Samasuwo: We'll kind of use... we'll always follow a very tightly controlled script, where we have clear themes and topics in that interview, so that when it comes to asking... potentially using AI to help us analyze that, instead of asking it to look through sort of a 2,000-word script, we can focus it on a particular area. So, look at all of the responses to the questions related to AI attitudes.

118

00:23:16.670 --> 00:23:27.290

Jesse Samasuwo: And instead of looking at, sort of, 10,000 words to try and analyze those, it's looking at a much smaller chunk, maybe 3 or 4,000 words, which increases the risk... which increases the quality of its outputs.

119

00:23:27.290 --> 00:23:46.229

Jesse Samasuwo: Really, really quickly, I know we were last just at time, a couple of principles and takeaways to think about. First of all, you are the expert. Brilliantly, I've forgotten to add the descriptive text here, so you can just see Lorem Ipsum. But important principles, you are the expert, and it's that expertise and contextual understanding that really makes the quality

120

00:23:46.230 --> 00:24:07.990

Jesse Samasuwo: That really makes a difference in high-quality analysis. Use tools that make your outputs auditable, so you don't want, instead of, as I said at the beginning, instead of just using AI to completely automate analytical outputs, use it to help you work through different stages of a process that others can query and look at and reproduce, and then think critically about AI's limitations in your specific context. So.

121

00:24:07.990 --> 00:24:13.929

Jesse Samasuwo: Again, using the example of, like, long context or short context windows, thinking carefully about how you can

122

00:24:13.930 --> 00:24:26.660

Jesse Samasuwo: chunk, or sort of, like, curate the integration that you want AI to work with. Think critically about how its limitations might apply in your context. A couple of pre-skills and, skills and capabilities. Curiosity.

123

00:24:26.990 --> 00:24:43.869

Jesse Samasuwo: try new tools, methods, even if in low-risk settings, so maybe it might be public data sets that maybe you're working with on an independent project, which you might then take back into the work context. If that's possible, I'm very mindful that people have different access to different tools, but that kind of curiosity to think about new data sets

124

00:24:43.870 --> 00:25:08.010

Jesse Samasuwo: create creativity to imagine where there might be data that can be used to provide richer insight. That's definitely something that everyone can bring to their teams and their organizations. You have a really, sort of, rich understanding on the ground of understanding what kinds of data flow through your team, or your organization, and thinking about ways in which that could help inform some decision-making or, sort of, service design. Again, looking back at that Facebook example.

125

00:25:08.010 --> 00:25:11.910

Jesse Samasuwo: earlier. And then, lastly, but perhaps most importantly, is just great communication.

126

00:25:11.910 --> 00:25:31.920

Jesse Samasuwo: So, data insight is nothing unless you can communicate it and help it, sort of inform impact. So, use your intuitive understanding of your audience, the people that you're working with, to find the most compelling way to deliver insights. That's a whirlwind tour. I spoke very quickly, so apologies if you didn't catch everything, but we can also do questions and pick it up in the community. Joel, back to you.

127

00:25:31.920 --> 00:25:44.899

Joel Smith: Fantastic. Thank you, Jesse, and sorry to rush you through it. As Jesse mentioned, we can pick up the conversation after. The recording of this event will be posted on the platform, so you can rewatch it if you missed anything.

128

00:25:44.900 --> 00:26:04.749

Joel Smith: Also, just worth briefly noting before we get into the Q&A, this is the first event in our AI Skills Labs series. We've got two more events coming up, they're live on the platform, so you can check them out. One in AI for email writing, and one in AI for policymaking prompts. So do check those out, they are just, going in the chat now.

129

00:26:04.750 --> 00:26:24.569

Joel Smith: So, on to the Q&A. We've only got a few minutes, so we'll try and hurry through as many questions as we can. First question for you, Jesse. One of our audience has asked, what are some methods to de-identify data points to be able to use AI on otherwise confidential data? Do you have any tips for just anonymizing datasets?

130

00:26:24.930 --> 00:26:30.729

Jesse Samasuwo: Yeah, I think all of those, obviously, you're working within a very sensitive context in government. I think some of those processes might

131

00:26:30.730 --> 00:26:54.449

Jesse Samasuwo: may or may not be suitable for using with AI. So, in our case, sometimes we have, user... private user identifier information from users. We absolutely cannot let that touch AI systems, so if we want to work with that, we might have to work with manual processes or sort of, like, safe environments, that are disconnected from internet to access that data. So, it really depends on your organizational context.

132

00:26:54.450 --> 00:27:06.429

Jesse Samasuwo: But there are some kind of offline tools and libraries, if you look at Python libraries, that help you kind of anonymize data. But again, I would really encourage you to make sure that you're kind of working within the boundaries of your organizational guidelines there.

133

00:27:06.730 --> 00:27:18.360

Joel Smith: Absolutely. Another question here from Sarah. Do you need to have some sort of programming background to be able to do this? Sarah says she's not a programmer, but would love to learn some basic LLM if possible.

134

00:27:19.130 --> 00:27:40.770

Jesse Samasuwo: Absolutely not, Sarah. I personally, myself, do not come from a technical background, and I think this is one of the great things about generative AI, is that it can help bridge the gap between non-technical and technical folks. So, some of these scripts, you can generate them just by natural language prompting, and having an understanding of some of the kind of analyses that you want to do, and asking AI to help you write scripts, and then play around in tools.

135

00:27:40.770 --> 00:27:47.329

Jesse Samasuwo: Personally, I use Google Colab, but there's others out there, I'm sure GitHub, you can download right onto your computer.

136

00:27:47.330 --> 00:28:01.150

Jesse Samasuwo: And just work with AI to kind of generate some scripts and see how they go. That's why I said, kind of, that curiosity and experimentation, download a public dataset which is not sensitive, and see, sort of, like, try out some different approaches.

137

00:28:01.810 --> 00:28:12.579

Joel Smith: Brilliant, lovely. I'm afraid we are running very close to time, so what we might do, we'll take these questions, we'll move it over into the AI and government community.

138

00:28:12.580 --> 00:28:32.080

Joel Smith: And, we'll try and get some of those answered in that space. I've seen there's a couple of questions around, kind of, Copilot as well. Worth noting, we've got an event on the horizon coming up in a couple of weeks, which is Hacks for Working with Copilot that is going to, kind of, meet all of your... all of your Copilot needs, so do check that out as well.

139

00:28:32.090 --> 00:28:48.439

Joel Smith: Okay, so we've had some incredible wisdom from Jesse, so I'd like everyone just to join me in giving Jesse a big virtual round of applause, and most importantly, thank you to you for the attendees, as always, for being so invested in your learning and being so engaged in the topic.

140

00:28:48.440 --> 00:28:59.370

Joel Smith: The team's gonna pull up a quick poll just to give us a rating on this event, how helpful you found it on a scale of 1 to 10, so please do vote on that, it helps us with shaping our content.

141

00:28:59.430 --> 00:29:15.709

Joel Smith: Before you go, and while you're responding to that, don't forget to register for the next AI Skills Lab events. If you're not available to join live, you can always register, and we'll send you the event recording after the session. So do get registering for those, we'd love to see you there.

142

00:29:15.720 --> 00:29:40.670

Jesse Samasuwo: Okay, that really is all we have time for today. It's been a real speedy session with a lot of information. As mentioned, the recording will go up on the platform tomorrow. We will send everybody an email with a link to that recording so you can watch it back in case you missed anything, and we will continue this conversation in the AI in government community. So, keep an eye out for that email coming your way, and we hope to see you all at another apolitical event very.

143

00:29:40.670 --> 00:29:44.049

Joel Smith: soon. Thank you very much, everybody. Goodbye.