

Transcript | What's Safe to Put Into AI (and What Isn't)

1

00:00:17.420 --> 00:00:29.780

Victoria Pallois: Hello, everybody. Hello and welcome. I'll just give it a couple of seconds for everybody to enter this virtual stage. The virtual room, we'll see.

2

00:00:30.330 --> 00:00:50.190

Victoria Pallois: Welcome to this webinar and masterclass on what's safe to put into AI and what isn't. My name is Victoria, and I'm an event producer at Apolitical. Today, we're lurking at a popular and recurring question. How do you know what's safe to put into AI?

3

00:00:50.650 --> 00:01:06.139

Victoria Pallois: Many of you might use tools like ChatGPT, Copilot, or Claude every day, but judging which pieces of data and information go into AI could create, and what those who can create a risk is not always straightforward.

4

00:01:06.530 --> 00:01:15.689

Victoria Pallois: Throughout the next 30 minutes, we'll learn about quick ways to judge what's risky, a few principles for deciding when AI should be involved.

5

00:01:15.820 --> 00:01:19.280

Victoria Pallois: And a process to check how sensitive your information is.

6

00:01:19.610 --> 00:01:23.800

Victoria Pallois: For those of you who may be joining us for the first time, welcome!

7

00:01:24.230 --> 00:01:32.770

Victoria Pallois: Apolitical is the world's largest network for government, used by more than half a million public servants and policymakers across the world.

8

00:01:32.890 --> 00:01:39.770

Victoria Pallois: We are a strategic partner to governments, helping them prepare their workforces for the future.

9

00:01:39.870 --> 00:01:51.150

Victoria Pallois: We achieve this by upskilling through short online courses and peer communities, as well as by building innovative learning tools designed for and with governments.

10

00:01:51.150 --> 00:02:01.430

Victoria Pallois: So, it'd be great to hear all about you, so please do post in the chat where you're joining us from, what organization you work for. So, I'm joining from London, UK.

11

00:02:01.700 --> 00:02:07.210

Victoria Pallois: So while you introduce yourselves, I'll get a few bits of housekeeping out of the way.

12

00:02:07.540 --> 00:02:15.760

Victoria Pallois: Firstly, I'll mention that this event is recorded for on-demand viewing on Apolitical's website, so please don't record it yourself.

13

00:02:16.340 --> 00:02:23.880

Victoria Pallois: We will also be using the recording to generate events insights, including with AI-powered software.

14

00:02:24.460 --> 00:02:33.959

Victoria Pallois: As this session is running on a Zoom webinar, we won't collect your audio or video data, but personal data shared in the chat may be captured.

15

00:02:34.640 --> 00:02:41.419

Victoria Pallois: Contact details for privacy concerns are available on the Apolitical homepage footer.

16

00:02:41.440 --> 00:02:48.759

Victoria Pallois: And closed captions are available in your own language, and can be toggled on or off in your screen settings.

17

00:02:48.760 --> 00:03:05.350

Victoria Pallois: just scroll up the chat to see that my colleague has posted the instructions. I can see so many introductions already, which is a great... let me have a look. We have people from Quebec, a lot of people from the UK, Manchester, London, Blackpool,

18

00:03:05.350 --> 00:03:24.060

Victoria Pallois: And Ecuador, oh my god, amazing. And the Bahamas, wow, and Dublin, wow, great. Okay, let me refocus again on the second piece of housekeeping. We would love to extend an invitation to join more than 30,000 of your peers in public service in our AI in government community.

19

00:03:24.310 --> 00:03:34.860

Victoria Pallois: This is an online space to share conversations, insights, ideas, and resources on the use of AI in government through connections and conversation with peers across the globe.

20

00:03:35.030 --> 00:03:49.279

Victoria Pallois: And it's open to all, whether you consider yourself an AI novice or an AI expert. So the link is in the resources tab of your screen, and my colleague also posted it in the chat, and I'd love to see your comments in there.

21

00:03:49.600 --> 00:04:07.710

Victoria Pallois: So let me have another look at all these... oh, people are coming from Carolina, Puerto Rico, South Wales, UK, Paris! Welcome, everybody! So many, so many, people from all over across the world, I love that. Well...

22

00:04:08.050 --> 00:04:25.299

Victoria Pallois: We, before we introduce our speaker, we're going to run one quick poll to get a sense of where you're joining, in today's discussion. So, the poll is going to pop up, and it's... the question is, when deciding whether to use AI, what's your biggest challenge?

23

00:04:26.400 --> 00:04:33.260

Victoria Pallois: You have a range of options to select. The first one is knowing what counts as sensitive information.

24

00:04:34.440 --> 00:04:38.540

Victoria Pallois: The second one is preparing data so it is safe to use.

25

00:04:39.850 --> 00:04:44.360

Victoria Pallois: The third one, getting sign-off or guidance from my organization.

26

00:04:45.890 --> 00:04:46.660

Victoria Pallois: Fourth?

27

00:04:46.940 --> 00:04:49.369

Victoria Pallois: I avoid using AI because of the risk.

28

00:04:50.170 --> 00:04:53.280

Victoria Pallois: And fifth, something else, please share in the chat.

29

00:04:54.000 --> 00:05:02.210

Victoria Pallois: So let me scroll down in the chat to see if anybody pasted or shared some... Insights...

30

00:05:03.740 --> 00:05:09.650

Victoria Pallois: I'll give you a couple of seconds more to vote. I can see so many of you have voted already.

31

00:05:10.900 --> 00:05:29.490

Victoria Pallois: Amazing. Okay, well, we're gonna close it now, and so, we have a winner, and the winner is knowing what counts as sensitive information. And a close second is preparing data so it is safe to use. Well, I really hope that today's session is going to help you answer these questions.

32

00:05:29.500 --> 00:05:49.400

Victoria Pallois: So thank you for taking part in this poll. Well now, introducing our amazing speaker, we're really fortunate to have Antonia Mochen joining us today. Antonia is a social scientist with a long experience in government. She is currently working in the European Commission Science Scientific Service.

33

00:05:49.400 --> 00:05:54.180

Victoria Pallois: the JRC, and a team that brings innovative methods to policymaking.

34

00:05:54.370 --> 00:06:00.609

Victoria Pallois: She is also part of a department-level team working on implementing AI within the JRC.

35

00:06:00.800 --> 00:06:18.759

Victoria Pallois: She's particularly interesting in how humans use AI and the changes it will make to our ways of working, learning, and collaborating. So, you can see, you can read Antonia's full bio by visiting the Resources tab at the bottom of your screen and heading to the Speakers section.

36

00:06:19.170 --> 00:06:23.110

Victoria Pallois: Well, Antonia, welcome. It's a pleasure to have you with us.

37

00:06:23.300 --> 00:06:46.499

Victoria Pallois: I'll... before I hand over to you, I'll just want to invite our speakers, well, no, our attendees, to submit your questions for Antonia in the Q&A tab that is at the bottom of your Zoom screen during Antonia's presentation, because we'll reserve a few minutes for Q&A at the end of Antonia's presentation. But Antonia, delighted to have you with us.

38

00:06:46.500 --> 00:06:47.799

Victoria Pallois: The floor is yours.

39

00:06:49.400 --> 00:06:55.300

Antonia Mochan: Thanks very much, Victoria. It's lovely to be here, and I hope I will go some way to

40

00:06:55.410 --> 00:06:59.880

Antonia Mochan: Maybe not answering your questions, that you've indicated, but at least

41

00:07:00.150 --> 00:07:07.590

Antonia Mochan: giving you some ways to think about how you might deal with them. So, I'm just gonna share my slides.

42

00:07:07.710 --> 00:07:09.630

Antonia Mochan: And let's get these up.

43

00:07:10.540 --> 00:07:12.070

Antonia Mochan: start here...

44

00:07:12.320 --> 00:07:22.349

Antonia Mochan: And start that up. So, the title is What's Safe to Put into AI and What Isn't? And I want to start with,

45

00:07:22.840 --> 00:07:27.419

Antonia Mochan: a word of caution. I am not you.

46

00:07:27.530 --> 00:07:38.850

Antonia Mochan: I don't work for the organization you work for, I don't do the job you do, I don't work the way you do. So, the things that I will say, I'm really trying to give you some

47

00:07:39.190 --> 00:07:53.109

Antonia Mochan: food for thought and some ideas, but please don't take this as kind of gospel, or I did it this way because Antonia told me to, because as with so many things about AI, it really does come down to how you use it, and it...

48

00:07:53.790 --> 00:07:57.639

Antonia Mochan: How it fits into your life and your work and your organization.

49

00:07:58.000 --> 00:08:03.960

Antonia Mochan: So think of me like an AI. Take what I say as an input, and examine it critically.

50

00:08:04.580 --> 00:08:11.939

Antonia Mochan: So, what I'm hoping to, share with you today are a few principles for working with AI, a couple of...

51

00:08:11.990 --> 00:08:28.860

Antonia Mochan: checklists for this issue of whether you should be putting data into, particularly an LLM. I mean, we use the word AI, but what we're talking about is generative AI systems and LLMs. And then I'm going to finish with, a prompting suggestion.

52

00:08:29.360 --> 00:08:33.669

Antonia Mochan: That I hope will help you to, to think about this topic.

53

00:08:34.380 --> 00:08:53.240

Antonia Mochan: So, these principles for human-AI collaboration were developed by colleagues in the Joint Research center who are doing a lot of work in this area of human-AI collaboration. And it all starts with complying with the rules. So, you need to know what your organizational rules are. These will be very different.

54

00:08:53.240 --> 00:09:07.290

Antonia Mochan: from a large organization like the Commission, or if you're in a small quango, or if you're in a startup that's working with government, you're going to have a very different starting point, so make sure you're aware of those rules.

55

00:09:08.210 --> 00:09:23.080

Antonia Mochan: And then you need to understand what AI can do, how it works, what it can't do, of course. And then you need to think about, just because it can do something, does that mean it should do it? Is generative AI the right answer to the thing that I'm trying to achieve?

56

00:09:23.910 --> 00:09:31.010

Antonia Mochan: you should give it a go. It's very much a test and iterate kind of process. Lots of people...

57

00:09:31.370 --> 00:09:34.940

Antonia Mochan: I know of, maybe not lots, but a few, they kind of...

58

00:09:35.760 --> 00:09:45.799

Antonia Mochan: go in once, give it a really bland prompt, don't like what happens, and go, it doesn't work! Well, you've got to kind of work with it a little bit, so do give it a probation period.

59

00:09:45.970 --> 00:09:50.359

Antonia Mochan: Learn to work with it, and keep your thinking active, and we'll come back to that point.

60

00:09:50.550 --> 00:09:54.770

Antonia Mochan: Stay in the driving seat. Verify, disclose, decide.

61

00:09:54.870 --> 00:10:10.089

Antonia Mochan: And finally, share and learn together, and that's the main thing we're doing here today, is sharing and learning together. And if there's any lesson, it is talk about these things with people who are doing them as well, because this is how we're going to get to grips with this technology.

62

00:10:10.760 --> 00:10:16.989

Antonia Mochan: As I said, though, I've got checklists, but I've actually got two checklists. One is about the data.

63

00:10:17.220 --> 00:10:31.849

Antonia Mochan: There are many shades of grey, I think, when you're thinking about the data that you would share with a generative AI system. There are some things that really don't need saying. Classified material, sensitive personal information.

64

00:10:32.040 --> 00:10:36.639

Antonia Mochan: And that will probably be, classified under a...

65

00:10:36.900 --> 00:10:44.660

Antonia Mochan: a national framework, or if you're in EU, something like GDPR is very clear about what is sensitive personal information.

66

00:10:46.490 --> 00:10:52.370

Antonia Mochan: Information that's not public, that your organization has not made public. That's really not something to put in.

67

00:10:52.650 --> 00:10:58.539

Antonia Mochan: pre-legislative information. So here you're getting into, you know, maybe slightly greyer areas.

68

00:10:58.680 --> 00:11:03.040

Antonia Mochan: But I think the real grey area is to remember that these are...

69

00:11:03.260 --> 00:11:14.190

Antonia Mochan: Connection and inference machines. They make connections, they decipher, they... they, make inferences. So do think about the context in the bigger picture.

70

00:11:14.960 --> 00:11:23.630

Antonia Mochan: it could be useful for analyzing my household budget, but do I want information about what I spend my money on to be available to a large company?

71

00:11:23.730 --> 00:11:30.189

Antonia Mochan: I could frame my request for input for a briefing note in very general terms.

72

00:11:30.450 --> 00:11:43.860

Antonia Mochan: But if I've already asked it to do something else work-related, it's going to make the connection between those two, and perhaps infer more than I'm comfortable with, or even less than I'm comfortable with. Make an assumption about what I do that isn't complete.

73

00:11:45.230 --> 00:11:54.740

Antonia Mochan: Another thing on the data side is that in some services and some description... subscription levels, you can turn off memory, so that's also quite useful.

74

00:11:54.960 --> 00:12:00.559

Antonia Mochan: But really, err on the side of caution. If you're not sure, probably don't do it.

75

00:12:00.750 --> 00:12:03.040

Antonia Mochan: And check your guidelines.

76

00:12:03.810 --> 00:12:07.330

Antonia Mochan: But I think even more interesting is the...

77

00:12:07.400 --> 00:12:17.659

Antonia Mochan: human checklist. And I've put a few things here. So, a colleague of mine also developed this slow, fast, slow framework for working with AI.

78

00:12:17.720 --> 00:12:32.579

Antonia Mochan: So, do some slow thinking first. Think about what you're asking the AI to do. Think about the things that you're putting in, the information that you're sharing, the context that you know and the machine doesn't know, the memories that you have that the machine doesn't have.

79

00:12:33.380 --> 00:12:39.819

Antonia Mochan: Then, if you've decided that it is the right thing, is to use AI, that's the fast bit, and it'll do its fast thinking.

80

00:12:40.060 --> 00:12:42.650

Antonia Mochan: And then you need to do the slow syncing again.

81

00:12:43.470 --> 00:12:54.810

Antonia Mochan: With what you got out, engage your critical thinking, engage your skepticism, even, about what it is. Apply the knowledge that you have. Ask someone else who maybe knows something about it that you don't know.

82

00:12:54.930 --> 00:12:58.439

Antonia Mochan: So that slow, fast, slow can be really useful.

83

00:12:59.640 --> 00:13:12.960

Antonia Mochan: Can, can't, should, shouldn't is another framework that I use. Just because AI can do something doesn't necessarily mean it should. And what are the implications of it if I put it in? So, it can generate

84

00:13:14.460 --> 00:13:22.979

Antonia Mochan: graphics to go with my, with my article, but does that mean my graphic designer colleague is sitting there doing nothing?

85

00:13:23.210 --> 00:13:33.319

Antonia Mochan: It can do an entry-level briefing quicker than my, than my new colleague, but then how is my new colleague going to learn if I don't ask them to do that briefing?

86

00:13:33.570 --> 00:13:36.109

Antonia Mochan: So that can, can't, should, shouldn't.

87

00:13:36.260 --> 00:13:42.210

Antonia Mochan: It's quite a good way of thinking before you use I also...

88

00:13:42.600 --> 00:13:52.699

Antonia Mochan: and I tested this with a colleague yesterday who was asking me exactly the same question, and I said to him, is this a... think about AI as having a conversation with someone.

89

00:13:53.380 --> 00:13:57.860

Antonia Mochan: Is this a conversation you'd be happy having in a public cafe?

90

00:13:59.100 --> 00:14:11.459

Antonia Mochan: Or, to put it another way, if you left the papers in the cafe, would you have a sinking feeling in your stomach, oh god, oh no, something terrible's gonna happen? Or will you be like, no one's gonna know what that is?

91

00:14:11.880 --> 00:14:16.629

Antonia Mochan: The CAFE test is quite a good way of thinking about it. When you're sat there, and you're about to

92

00:14:17.000 --> 00:14:24.089

Antonia Mochan: Put something there, think, would I have this conversation in public? And if you wouldn't, then you probably shouldn't be having it with an AI.

93

00:14:25.510 --> 00:14:31.710

Antonia Mochan: And then the last one is using AI to prompt the human, and that's what I'm going to show you now.

94

00:14:32.750 --> 00:14:44.949

Antonia Mochan: So, I use this quite a lot. So, instead of me asking the AI to do something, I ask it to help me think through what it is I'm trying to do. And I have this on quite a lot

95

00:14:45.170 --> 00:14:51.339

Antonia Mochan: trying to build a prompt so that I have a really good prompt, but also to walk me through where I should use it.

96

00:14:51.500 --> 00:14:53.510

Antonia Mochan: And I used that.

97

00:14:53.620 --> 00:15:04.959

Antonia Mochan: to build a prompt to share with you today. We will share the slides with you, so you will have this afterwards, so don't worry about kind of trying to scribble it down now, but just, take a look.

98

00:15:05.590 --> 00:15:12.080

Antonia Mochan: So, what I asked the AI to do was to give me a prompt that I could use

99

00:15:12.560 --> 00:15:15.620

Antonia Mochan: to think through whether I should be using something.

100

00:15:15.800 --> 00:15:16.640

Antonia Mochan: So...

101

00:15:17.130 --> 00:15:24.230

Antonia Mochan: You're acting as a cautious prompt engineering assistant. Help me check whether a task I want an AI to do involves sensitive information.

102

00:15:25.000 --> 00:15:28.049

Antonia Mochan: The bits in highlight are things that you would

103

00:15:28.380 --> 00:15:30.570

Antonia Mochan: Be able to, or you should.

104

00:15:30.890 --> 00:15:45.079

Antonia Mochan: put in the information relative to your organization, your systems, what counts as confidential, etc. there. So these are the more customizable bits. But basically, I'm asking the AI to walk through me... through with me in steps.

105

00:15:45.240 --> 00:15:48.300

Antonia Mochan: whether this is something that I should use AI for.

106

00:15:49.830 --> 00:15:58.709

Antonia Mochan: And I'm gonna show you some screenshots of where I asked it to do this. I thought a live demo was perhaps not the best idea, but...

107

00:15:58.790 --> 00:16:11.850

Antonia Mochan: I promise you, I just did this when I put the slides together a couple of days ago, and you can try it yourself and see what comes out. And what I decided to ask it to do was to help me put together a budget.

108

00:16:12.350 --> 00:16:16.050

Antonia Mochan: So, what's... so, analyze data.

109

00:16:16.950 --> 00:16:30.380

Antonia Mochan: what kind of data is it? So it's budget data for next year, it's currently internal only. I want to analyze it to see where we could make efficiencies, for example, where projects are overlapping, and to understand how much we're spending in different headings.

110

00:16:31.360 --> 00:16:40.570

Antonia Mochan: So I put that in, and it said to me, okay, two more questions. Does it name people? Does it name external people? Has it been publicly announced?

111

00:16:40.860 --> 00:16:46.289

Antonia Mochan: And I said, yeah, the names of the project leaders are included, and it's still at the draft stage.

112

00:16:47.120 --> 00:16:48.750

Antonia Mochan: And it said to me.

113

00:16:48.870 --> 00:17:06.989

Antonia Mochan: This is really likely sensitive, and I recommend avoiding general external AI tools. And it gave me a whole bunch of reasons as to why, and it gave me some suggestions as to how that might be something that it would be more... more,

114

00:17:07.310 --> 00:17:12.990

Antonia Mochan: responsible to use, if you were going to use AI, how to make it less sensitive in the data.

115

00:17:13.589 --> 00:17:24.559

Antonia Mochan: So, strip out names, use an internal tool if you have one, generalize the figures and the names, and then here, one honest caveat is your critical friend.

116

00:17:24.859 --> 00:17:32.810

Antonia Mochan: It could still be distinctive enough that someone could reverse engineer it. So, worth a gut check on that, not just a mechanical find and replace.

117

00:17:33.400 --> 00:17:45.270

Antonia Mochan: So, for me, it's given me a bunch of things that I need to think about. It's not telling me what to do, it's not replacing my judgment, but it's just prompting me to make some decisions myself.

118

00:17:45.530 --> 00:17:55.870

Antonia Mochan: I saw a presentation about AI by a prominent neuroscientist, and he said one of the things is that decisions cost us energy, and

119

00:17:55.870 --> 00:18:07.500

Antonia Mochan: they have implications, and the AI, there is no implications for the decision it makes. So we really need to sit in the driving seat, coming back to the principles, but we can use AI to help us do that.

120

00:18:07.690 --> 00:18:16.140

Antonia Mochan: So I would say, when we're thinking about putting data in, we need to measure twice and cut once, as they say in, in carpentry.

121

00:18:16.290 --> 00:18:23.550

Antonia Mochan: Take the time to think through the decision you're making, so that you're not going to do something that, you might regret later.

122

00:18:24.410 --> 00:18:30.029

Antonia Mochan: So, that's just a quick run through some thoughts, and happy to move on to some Q&A.

123

00:18:31.350 --> 00:18:54.129

Victoria Pallois: Amazing! Thank you so much, Antonia, and I think the... your cafe test, I think it depends maybe on the generation, but I do more of a social media test, you know, would I paste this information on social media for everybody to read? Maybe not sometimes. But I guess we all have our own, like, references.

124

00:18:54.370 --> 00:19:04.800

Victoria Pallois: So we'll, we'll make sure to have all the slides available on the event page after the event, so that you can have a copy of this prompt, and you can also read it,

125

00:19:04.800 --> 00:19:14.679

Victoria Pallois: In your own time. We have some amazing questions, actually, already in Q&A, so, but please do drop... continue dropping them in the Q&A tab.

126

00:19:15.220 --> 00:19:21.600

Victoria Pallois: Just before we dive into your questions, I wanted to let you know that we've also curated a great follow-on resource.

127

00:19:21.660 --> 00:19:36.529

Victoria Pallois: On today's topic. It's a lesson that covers certain users, how certain users of AI in the public sector come with higher risk of harm, and teaches you how to identify them and understand the impact.

128

00:19:36.530 --> 00:19:43.709

Victoria Pallois: So, to receive your link to the short lesson, go to the AI and Government community after the event, and leave a

129

00:19:43.710 --> 00:19:49.790

Victoria Pallois: an answer to my question, what do you know about your organization's AI policy?

130

00:19:49.790 --> 00:20:01.099

Victoria Pallois: on the post pinned at the top of the page. My colleague is also sending a link to that post in the chat, and we'll make sure to send the resources to you after the event.

131

00:20:01.970 --> 00:20:05.209

Victoria Pallois: So, on to the Q&A.

132

00:20:05.470 --> 00:20:26.790

Victoria Pallois: I'm gonna start with what I would say an easy question, but they're all a bit really tricky, really. You recommended, Antonia to use an internal tool. Do you mean using an LLM developed by your own organization, or using an enterprise license of, say, like, Copilot, ChatGPT, etc?

133

00:20:27.790 --> 00:20:31.319

Antonia Mochan: When I say internal tool, I tend to mean your organization's,

134

00:20:31.680 --> 00:20:43.710

Antonia Mochan: own LLM, because there's going to be a set of guardrails, etc, that are relevant to your organization, and probably going to be helping... and the guidelines will be

135

00:20:43.890 --> 00:20:48.410

Antonia Mochan: customized to fit that LLM. So,

136

00:20:49.330 --> 00:21:07.020

Antonia Mochan: So yeah, I think if something can be done there, it should be, and if something is sensitive, then even if you don't get as good, an output as you might in a paid-for external model, then really, you need to use the internal one, because that's the

137

00:21:07.250 --> 00:21:22.859

Antonia Mochan: where you can safely use sensitive data. But really get familiar with your organization's guidelines, because that is gonna... and as I said at the beginning, I don't know your organization, I don't know your rules, I don't know your role, so that's really got to be your starting point.

138

00:21:23.530 --> 00:21:36.470

Antonia Mochan: And what you can do, of course, is customize the prompt that I gave you, or generate one for your organization, where you can put those in very specifically. So, I actually built that with a skill that

139

00:21:36.470 --> 00:21:46.819

Antonia Mochan: knows what my organization's rules are, and can really help me very specifically, make a decision about what tool I should use for the piece of work that I'm doing.

140

00:21:47.940 --> 00:22:01.850

Victoria Pallois: Thank you. Well, we have two most upvoted questions that I'm gonna ask together. So if you stay within your organization's tenant or, like, tool, what is the principal risk, then?

141

00:22:03.510 --> 00:22:09.579

Antonia Mochan: Well, there isn't particularly a risk, I would say. I mean, I would hope that there isn't a risk of that. The risk

142

00:22:10.100 --> 00:22:13.029

Antonia Mochan: a data risk. The risk is...

143

00:22:13.870 --> 00:22:20.380

Antonia Mochan: All of the implications of us using... of us outsourcing our thinking to something that isn't us.

144

00:22:20.600 --> 00:22:32.340

Antonia Mochan: And so that's why we talk about the human-AI collaboration, and that's why we encourage you not to think about just what AI can do, but whether it should do it, because there are all these implications, such as

145

00:22:32.470 --> 00:22:48.209

Antonia Mochan: Data is one of them, but impact on colleagues, impact on your own critical thinking. There's growing numbers of studies now showing that we are doing this sort of cognitive offloading to machines. So that's what the risk is.

146

00:22:48.210 --> 00:23:02.719

Antonia Mochan: But of course, there are benefits to be had if you use this tool wisely, and I think that's the really important thing, is to put those... those stage gates into your own thinking, and that's why I find the slow, fast, slow such a useful framework.

147

00:23:02.920 --> 00:23:07.870

Antonia Mochan: I do a bunch of thinking, and I make a decision about what I ask the AI to do.

148

00:23:08.410 --> 00:23:13.720

Antonia Mochan: having thought through the implications of asking it to do that, and then I don't just

149

00:23:13.970 --> 00:23:24.770

Antonia Mochan: suck up what it gives me back, but I examine it critically. And I know, working in government, it can be really tempting when you... the pressure is on, and something needs to be done really quickly.

150

00:23:24.770 --> 00:23:34.600

Antonia Mochan: And, you know, that is something that happens everywhere. I understand the temptation to do that, but it's just about taking a little moment to think, okay, but what are the longer-term implications of that?

151

00:23:35.790 --> 00:23:52.049

Victoria Pallois: I think that's an... that's a very interesting answer, and we actually have a question that kind of feeds into that. Given, like, how AI... the usage of AI is increasing so much, and that it's, like, its capabilities is, you know, expanding day to day.

152

00:23:52.090 --> 00:23:57.120

Victoria Pallois: How realistic do you think it is to expect a human in the loop?

153

00:23:57.170 --> 00:24:05.590

Victoria Pallois: to be kept in everything, in all of AI outputs. Do you think that at some point, just the human is going to be kicked out of this loop?

154

00:24:06.870 --> 00:24:18.950

Antonia Mochan: I mean, I don't have a crystal ball, but I think we really do have a lot of agency there. I think it depends also slightly what you're asking it to do, so if it's... I work in research, so I have colleagues who are, kind of.

155

00:24:19.130 --> 00:24:27.320

Antonia Mochan: Running experiments, you know, analyzing data, updating, dashboards or something, which is, you know.

156

00:24:27.910 --> 00:24:38.379

Antonia Mochan: there's perhaps not the level of involvement that's necessary in the more qualitative kind of work that I'm doing, where, you know, if you're briefing a minister.

157

00:24:38.490 --> 00:24:45.950

Antonia Mochan: you do... you want to make sure that you're in the loop, that that minister is getting genuine information.

158

00:24:46.540 --> 00:24:48.279

Antonia Mochan: So, so yeah, I, I...

159

00:24:48.490 --> 00:24:56.630

Antonia Mochan: I understand the question, but I think, you know, in a way, it's a kind of existential question we have to keep.

160

00:24:56.890 --> 00:25:03.289

Antonia Mochan: we have to keep in the loop, because, you know, that's... that's who we are and what we're doing.

161

00:25:03.430 --> 00:25:12.479

Antonia Mochan: That made me think of something else I wanted to say, but it's just gone. Yes, that was it. I think, when we think about the benefits, there are...

162

00:25:12.700 --> 00:25:31.759

Antonia Mochan: there are some ways in which... I work in a multilingual organization, so, we... how we have three working languages, we have 24 languages in the... in the union, we have a lot of people that... working with me, who are writing documents for whom English, predominant working language is maybe their third or fourth language. So.

163

00:25:31.850 --> 00:25:44.940

Antonia Mochan: LLMs can be incredibly helpful there in polishing text, in making it better for translation, for example, in smoothing the way that people work together, and that's absolutely fine.

164

00:25:45.070 --> 00:25:53.659

Antonia Mochan: The question comes, then, on the more critical thinking and the more judgment-oriented work, and I think there it's incredibly important that we do stay in control.

165

00:25:55.460 --> 00:25:59.730

Victoria Pallois: Yeah, I do agree with you that sometimes it's...

166

00:25:59.730 --> 00:26:19.569

Victoria Pallois: makes your work look so easy, like it's translated this message or polished your email so well that you just want to copy and paste and send, but it does make a mistake, especially if you made a little mistake, small mistake in your input, it comes back... it comes out, like, tenfold. So always proof... proofread, whatever.

167

00:26:19.570 --> 00:26:27.299

Antonia Mochan: Yeah, from the beginning, I have treated, generative AI as a not-that-good intern.

168

00:26:27.510 --> 00:26:43.969

Antonia Mochan: So, it's really useful for doing stuff, but you can't trust the output enough to send it without checking it. And even as they get better, I can't say that I've changed that mental model. I still think of it as a kind of slightly crap intern.

169

00:26:44.340 --> 00:26:45.300

Victoria Pallois: Yeah.

170

00:26:45.320 --> 00:26:53.289

Victoria Pallois: Okay, well, we've talked a lot about, like, the content, the document content, like, words, emails, all of that.

171

00:26:53.290 --> 00:27:10.519

Victoria Pallois: But what about the metadata of documents, such as, like, usernames, IP addresses, timestamps, or all of that? Should that be also considered as, you know, important data that should be monitored before submitting it into AI?

172

00:27:11.210 --> 00:27:21.080

Antonia Mochan: I would say yes, largely because of the point I made about inference and connection and that sort of thing. And, so you could be putting a very...

173

00:27:21.100 --> 00:27:38.189

Antonia Mochan: bland... you could think, I'm putting this in here, and it's... you know, nobody would know that I'm talking about my work, but you're doing it from a work IP address, and it knows where you work. So, you do need to be... you need to be thinking not just about what's going in, but that broader context.

174

00:27:38.300 --> 00:27:42.500

Antonia Mochan: Absolutely. And I think... I think that's where, again, I mean.

175

00:27:43.110 --> 00:27:56.109

Antonia Mochan: for the older generation, the cafe thing is, like, would you sit in a cafe with your... with your work badge talking about your work, you know? It's not just what you say, it's... it's...

176

00:27:56.250 --> 00:28:04.780

Antonia Mochan: the badge that you're showing, it's the... it's the case... the attache case you've got that's got the company logo on it. So there's all of this contextual stuff.

177

00:28:04.880 --> 00:28:13.559

Antonia Mochan: That you can be putting into that machine without really knowing about it. And in fact, we could update the prompt to include those kind of things, indeed.

178

00:28:15.020 --> 00:28:22.829

Victoria Pallois: Okay, we have time for one last question, and I'm kind of torn between so many questions, but I think...

179

00:28:23.510 --> 00:28:36.120

Victoria Pallois: On the top of your mind, Antonia, do you have an example of maybe, like, a big mistake somebody made in this area of submitting sensitive information into AI, and what we can learn from this mistake?

180

00:28:38.210 --> 00:28:50.010

Antonia Mochan: I don't think I can think off the top of my head of, like, you know, one of those urban stories that people tell each other, oh my god, you'll never guess what X did, but I think,

181

00:28:50.530 --> 00:28:58.189

Antonia Mochan: And I think that's kind of the problem, in that it isn't a one-off shot, something terrible, but it's an insidious kind of thing. It's...

182

00:28:58.250 --> 00:29:12.970

Antonia Mochan: We don't see it, it's incremental, and that's why we need to be doubly careful. Because if it was a question of, I put it all in there and then it appeared, you know, it appeared somewhere and it was all terrible, then we would have a cautionary tale that we would react to better as a human.

183

00:29:12.970 --> 00:29:29.859

Antonia Mochan: But... but in a way, what we're doing is we're boiling the frog. We're just letting these things increment and increment, and then suddenly we'll be... we'll be boiled and we won't know. So that's why I think keeping that critical human judgment before you do anything is so important. The best way you can

184

00:29:29.910 --> 00:29:32.759

Antonia Mochan: stop something bad happening is by

185

00:29:33.540 --> 00:29:36.599

Antonia Mochan: Reflexively thinking about something before you do it.

186

00:29:37.570 --> 00:29:40.670

Victoria Pallois: Yes. Measure twice, cut once.

187

00:29:40.670 --> 00:30:01.649

Victoria Pallois: We're all carpenters at the end of this event. Okay, well, I think we unfortunately have to stop it there, but we've had some really incredible wisdom from you, Antonia, so please,

I'll ask... I'd love to ask you all in joining to send all the, as many emojis as you can to give Antonia a virtual round of applause.

188

00:30:01.650 --> 00:30:11.509

Victoria Pallois: And most importantly, thanks to you, the audience, for investing the time and being so engaged in this topic. I can see a lot of comments that you want a part 2 and 3.

189

00:30:11.510 --> 00:30:23.079

Victoria Pallois: You know, as the events team here, we want to hear your ideas, so please drop in the community what you want to hear about in Part 2 and 3, you know, what are the questions that you still have?

190

00:30:23.130 --> 00:30:32.859

Victoria Pallois: So the team is going to pull up a quick poll to see how you felt about today, and this really helps us improve our content, and we really appreciate your responses.

191

00:30:32.860 --> 00:30:50.359

Victoria Pallois: So, before you go, and while you're responding to that poll, I want to share a few more events that you can register for, so you can apply what you've learned today. Remember that if you're not available to join live, please do register so that we can send you the recording directly, and you can catch up anytime.

192

00:30:50.510 --> 00:30:57.060

Victoria Pallois: So, in two weeks, we have our next AI Skills Lab edition on AI for Better Briefings.

193

00:30:57.060 --> 00:31:12.879

Victoria Pallois: And you can catch up on all the past editions of AI Skills Lab in the AI Skills Lab Hub. So you can share the recordings with your colleagues, with, you know, if you've remembered something cool, but you don't remember where, we have highlights everywhere.

194

00:31:12.880 --> 00:31:37.489

Victoria Pallois: So, that's all we have time for today, so we hope, we really hope you found it really helpful. We want to say a huge thank you to Antonia, to everyone who's made today happen, and to you for taking the time out of your busy days. Keep an eye out for the email coming your way with the resources, the recording. Don't forget to post under my post in the community, and we hope to see you at another Apolitical event soon.

195

00:31:37.590 --> 00:31:39.949

Victoria Pallois: Have a great day! Bye-bye!

apolitic.com