**ZOOM WEBVTT Transcript** (please note that Zoom transcript is not exact, Zoom types what it thinks it hears)

**Dr. Didong Li, University of North Carolina at Chapel Hill, NISS 2024 Writing Workshop at JSM on ChatGPT**

Sunday, August 4, 2024

1

00:00:01.460 --> 00:00:03.140

Didong Li: Good morning, everyone.

2

00:00:03.760 --> 00:00:12.540

Didong Li: it's great to see everybody in person. Finally, I'm piano Mo. I hope you had a nice trip to Portland from wherever you were in the world.

3

00:00:13.128 --> 00:00:18.000

Didong Li: So we're actually going to have the last session, last segment of our workshop.

4

00:00:18.230 --> 00:00:19.180

Didong Li: And

5

00:00:19.950 --> 00:00:27.290

Didong Li: our 1st speaker today is Dr. Duda only, and he is going to give us an introduction to Chat Gpt

6

00:00:27.520 --> 00:00:31.970

Didong Li: before we get started. I'd like to give a very short introduction to Dr. Lee.

7

00:00:34.090 --> 00:00:39.950

Didong Li: Dr. Dinoli is an assistant professor in the department of Biostatistics at Unc. Chapel Health.

8

00:00:40.100 --> 00:00:48.040

Didong Li: He, his research focus is statistical methods, development for robust inference of complex and high dimensional data.

9

00:00:48.240 --> 00:01:13.280

Didong Li: specifically covering manifold learning, downparametric Bayesian inference, information, geometry and spatial statistics. He has applied these methods to electronic health care record data and large scale genetic and survey data and single cell Rna. Sequencing data. In 2,019, Dr. Lee received the inaugural Ims. Lawrence D. Brown, Phd. Student award by Ims

10

00:01:13.430 --> 00:01:14.389

Didong Li: or 2 weeks.

11

00:01:17.440 --> 00:01:20.230

Didong Li: Thanks for the American introduction

12

00:01:20.840 --> 00:01:22.952

Didong Li: and the invitation. So

13

00:01:24.010 --> 00:01:29.470

Didong Li: Oh, morning, every1. 0, I'm sorry that I missed the previous 2 sessions.

14

00:01:30.010 --> 00:01:30.690

Didong Li: Nope.

15

00:01:30.980 --> 00:01:32.460

Didong Li: let's see the the box. And

16

00:01:36.770 --> 00:01:38.350

Didong Li: sorry, Buddy, we're trying to record it.

17

00:01:39.470 --> 00:01:41.289

Didong Li: Yeah, those.

18

00:01:43.120 --> 00:01:44.079

Didong Li: It doesn't disappear.

19

00:01:46.590 --> 00:01:47.450

Didong Li: Okay?

20

00:01:47.700 --> 00:01:48.490

Didong Li: Great.

21

00:01:52.160 --> 00:02:00.359

Didong Li: Okay? Okay? No. Problem. Yeah, I I hope you enjoy the the 2 online sessions. I I'm sorry I missed that.

22

00:02:01.600 --> 00:02:05.239

Didong Li: So today we will discuss something of

23

00:02:05.270 --> 00:02:09.700

Didong Li: probably very different. It's not about paper writing or grant writing.

24

00:02:09.759 --> 00:02:14.570

Didong Li: It's about a new tool chat bt, possibly known to everyone.

25

00:02:15.320 --> 00:02:19.889

Didong Li: So before we start, I would like to do a quick survey.

26

00:02:19.900 --> 00:02:25.229

Didong Li: So how many of you have heard about Gpt, please?

27

00:02:27.270 --> 00:02:36.739

Didong Li: Almost almost everyone. So this is, this is pretty different from the population. Survey results. So here's the table from a

28

00:02:37.390 --> 00:02:39.730

Didong Li: AI report by Stanford.

29

00:02:40.350 --> 00:02:42.629

Didong Li: Probably it's too small. Can I see the numbers?

30

00:02:43.130 --> 00:02:51.530

Didong Li: Nope, okay, I'll I'll read some key numbers here. So this array asks people whether you have. You are aware of

31

00:02:52.520 --> 00:02:54.760

Didong Li: okay? And so

32

00:02:55.210 --> 00:02:59.600

Didong Li: globally, about 63% people know Tai, Gbt.

33

00:03:00.130 --> 00:03:17.280

Didong Li: and for example, in United States number is 54%, which is a little bit surprising, because Hgbt was born in this country. But it's not as as well known as other countries or globally. For example, in India, about 82% people are aware of.

34

00:03:18.723 --> 00:03:22.529

Didong Li: So what impressed me was when I,

35

00:03:22.690 --> 00:03:36.199

Didong Li: when I was traveling to Hawaii last summer, I suppose in July I was attending a machine learning conference called Icml. It's 1 of the top Machine Conference. A lot of other people working on this live language model. In that conference

36

00:03:36.730 --> 00:03:47.980

Didong Li: I arrived at Airport I uber to the convention center, and the Uber driver asked me like. Why, you why you come to Hawaii, I said. I'm I'm here for a conference.

37

00:03:47.990 --> 00:03:51.100

Didong Li: and the Uber. The Uber driver asked me, what conference

38

00:03:51.410 --> 00:03:54.479

Didong Li: is it for business or for something else.

39

00:03:54.580 --> 00:03:58.209

Didong Li: I said, this is a more like a research conference for machine learning.

40

00:03:58.230 --> 00:04:06.389

Didong Li: Attendees are mainly like students, professors from academia and some tech industry. People like

41

00:04:06.560 --> 00:04:08.329

Didong Li: like Google, for example.

42

00:04:08.980 --> 00:04:23.140

Didong Li: So I, this conversation happens very often to me when I go to a a conference. You know, the Uber driver asked me why you come here? I just said, Okay, I came. I came here for a Stats conference or a machine learning conference the yearly. Say, oh, cool! That's it.

43

00:04:23.390 --> 00:04:32.109

Didong Li: But this time, after I I told the driver I'm here for machine learning conference, driver said, Oh, I know machine learning. I, I'm using chat Gpt.

44

00:04:33.020 --> 00:04:40.850

Didong Li: So this was the really the 1st time to me. And then I was aware of this how popular technology was at that time.

45

00:04:40.980 --> 00:04:49.660

Didong Li: That was in the summer of 2023. I think that was tragedy 3.5. By that time it's it was was not distributed for you.

46

00:04:50.160 --> 00:04:51.010

Didong Li: Oh.

47

00:04:51.560 --> 00:05:08.139

Didong Li: but now the the latest version is Gp. 4. 0, which is even much stronger than than that one. So this array was done in 2023, although the report it's came out the early this year, so I believe the number is even higher. If you do the same survey this year.

48

00:05:08.740 --> 00:05:13.610

Didong Li: and so let's let's do another quick survey. How many of you have used? Chat gpt

49

00:05:17.040 --> 00:05:17.940

Didong Li: again

50

00:05:18.760 --> 00:05:21.570

Didong Li: almost everywhere, like at least 80%.

51

00:05:22.560 --> 00:05:32.139

Didong Li: And so this is the number from the report. Let me read some numbers here. So globally, about 17% people use high Gbt daily

52

00:05:32.470 --> 00:05:39.070

Didong Li: and about half of the global population use chat at least weekly.

53

00:05:39.350 --> 00:05:45.059

Didong Li: So that's the global average. And if you look at United States, 18% for daily users

54

00:05:45.160 --> 00:05:47.880

Didong Li: and 27% for weekly users.

55

00:05:48.480 --> 00:06:03.909

Didong Li: And if you also count the users who use it monthly, the number can be 88% in China, or maybe 80, 85% in India. So ours is pretty popular internationally.

56

00:06:04.230 --> 00:06:06.209

Didong Li: Maybe that's why

57

00:06:06.950 --> 00:06:11.910

Didong Li: The organizer is is considering to to discuss this in the in the writing workshop.

58

00:06:13.960 --> 00:06:25.329

Didong Li: So when when Dan reached out to me like, Hey, do you know, can you? Can you present the A can you give a presentation about how to use Hpt in your research, especially in writing.

59

00:06:26.000 --> 00:06:35.010

Didong Li: I said, Okay, because I use it every day. But when I prepared for the slice, I just realized that how challenging this presentation is.

60

00:06:35.440 --> 00:06:38.760

Didong Li: So. Imagine you are giving a research talk.

61

00:06:38.940 --> 00:06:45.519

Didong Li: You know you prepare for your slice. You know what's going on on the next slide. You can prepare the transition. You know what's happening.

62

00:06:45.590 --> 00:06:47.360

Didong Li: Every single slide

63

00:06:47.400 --> 00:06:49.630

Didong Li: and the output is deterministic.

64

00:06:50.440 --> 00:07:03.090

Didong Li: But the presentation for today is is really random, because you know the name chat. Gpt. G means generative. That means the result from Gpt is generated from a probability distribution

65

00:07:03.330 --> 00:07:05.180

Didong Li: it? It gives you the

66

00:07:05.230 --> 00:07:17.700

Didong Li: answer with the highest probability. So that means, even if you ask the same question, on the same laptop, with the same account, you may have very different answers. I guess, if you have tried this, you may notice this.

67

00:07:17.970 --> 00:07:27.829

Didong Li: And now, so today we will play together on some simple questions, and it's very likely that we have very different results on our laptops and our different accounts.

68

00:07:28.380 --> 00:07:35.710

Didong Li: So it's it's it's not under my control. So I may look like that after this. This presentation.

69

00:07:35.730 --> 00:07:53.740

Didong Li: But I'll I'll I'll try my best. But I apologize, you know, you know, at once if I mess up anything or I get stuck somewhere. I prepared some sample answers by a screenshot. But I'll I'll do the questions with you together today again. So I bet that hasn't feel very different.

70

00:07:54.340 --> 00:07:55.550

Didong Li: so let's see.

71

00:07:56.840 --> 00:08:01.809

Didong Li: So here's a brief overview of the presentation. It will be pretty long.

72

00:08:02.050 --> 00:08:13.149

Didong Li: I'll I'll I'll 1st do a very brief introduction to Tehbt. Not really technical, because this is a writing workshop. It's not a research workshop on large language models.

73

00:08:14.060 --> 00:08:23.250

Didong Li: And then we have a couple of sections. One is about reading where we will, we will use Gpt. To summarize. A research paper is a published paper.

74

00:08:24.220 --> 00:08:32.380

Didong Li: Then, in terms of writing, we will use Gpt. To write an outline for a manuscript and do some grammar correction.

75

00:08:32.919 --> 00:08:37.839

Didong Li: So I will show you some sample text so that we can do the grammar correction together.

76

00:08:38.140 --> 00:08:40.149

Didong Li: So that'd be on the same page.

77

00:08:40.260 --> 00:08:43.729

Didong Li: The next chapter is about latex generation.

78

00:08:43.950 --> 00:08:58.730

Didong Li: because when we write paper sometimes the red papers in Latex, and you know Latex is not very smart in handling tables and figures, especially say if the figure has multiple panels with different size misaligned.

79

00:08:58.770 --> 00:09:05.040

Didong Li: it's pretty challenging to to deal with the later code or the same for complex tables.

80

00:09:05.810 --> 00:09:17.140

Didong Li: And the last main chapter view about coding, which is a key component in statistics. So we will explain us. Gpt, to explain our code

81

00:09:17.610 --> 00:09:18.660

Didong Li: and so

82

00:09:18.750 --> 00:09:33.230

Didong Li: generate our code for simple tasks. For example, linear regression. Maybe our should be to debug our no debug is painful in your code. Maybe we spend like 1Â h coding. We spend a hundred hours debugging.

83

00:09:33.660 --> 00:09:48.340

Didong Li: Oh, then, like so this is the last one. Did you translate our code to python. This is also pretty important. If you collaborate with someone else who's more comfortable with another language, maybe not python, maybe some something else.

84

00:09:49.280 --> 00:09:50.330

Didong Li: Oh, then

85

00:09:50.350 --> 00:10:01.309

Didong Li: we will discuss some limitations and some extensions. Not only chat Tp itself, maybe there are some other other large language model tools to help us with this writing.

86

00:10:02.880 --> 00:10:10.430

Didong Li: Okay, so let's start. Oh, here I have a very brief introduction over to the the history of tech. Gpt.

87

00:10:10.470 --> 00:10:16.440

Didong Li: so what does Gbt. Mean it is is the full name is generative, pre-trained transformer

88

00:10:17.715 --> 00:10:36.350

Didong Li: which is a special type of large language model. She also called this AI. Sometimes. So the the 3 words are actually all important here. So generative means is a generative model, you know, if you ask a question, it will generate an answer from some distribution learned from the model.

89

00:10:37.340 --> 00:10:52.159

Didong Li: and the second word pre trained means. The model has been trained already. There we will see how big the model is, there's no way for us to retrain it on our laptop. So you have to use the pre trained model. Somehow

90

00:10:52.850 --> 00:11:03.440

Didong Li: the 3rd World Transformer shows the main architecture of the Gpt or the large language model, which is still the most popular architecture for life, language models.

91

00:11:04.710 --> 00:11:16.530

Didong Li: and the the architecture of transformer is really from this 2017 new ribs paper. Most authors are from Google at that time ran from Toronto.

92

00:11:16.840 --> 00:11:33.700

Didong Li: and the paper title is Attention is all you need, and maybe you have seen a similar format in in paper titles. This is probably the 1st one you will see. Something is all you need, like linear regression is all you need. You can. You can do something like that. This is very popular nowadays, at least in machine learning.

93

00:11:34.430 --> 00:11:36.180

Didong Li: They're the.

94

00:11:36.370 --> 00:11:55.969

Didong Li: They studied a mechanism in deep learning, called attention, which was invited a couple of years before this paper, and they make it more scalable. They do some engineer tricks to make it highly scalable to much bigger training data sets. That's how they make this Gpu work. And it was trained basically on the entire Internet.

95

00:11:56.960 --> 00:12:10.469

Didong Li: And the 1st Gpt. Was really from this was born in this report is is not published. As far as I know. It's called a technical report in 2,018 by 4 authors from Openai.

96

00:12:11.278 --> 00:12:16.210

Didong Li: The title is Improving language understanding by generating, pre-training.

97

00:12:16.260 --> 00:12:22.889

Didong Li: and so the the tool they released in this technical report is now known as Gpt. One.

98

00:12:23.510 --> 00:12:30.829

Didong Li: So here I copy this table from Wikipedia you can see the 1st row is the 1st Gpg model ever

99

00:12:31.080 --> 00:12:32.660

Didong Li: on some key numbers.

100

00:12:32.930 --> 00:12:38.490

Didong Li: There are 12 levels and 12 highly transformer decoder. It's a transformer, just a transformer.

101

00:12:38.710 --> 00:12:46.599

Didong Li: Of these are a hundred 70 million parameters. This is probably much bigger than any statistical model you have seen.

102

00:12:47.216 --> 00:13:01.130

Didong Li: At least in my career. This is the largest model it was trained on 7,000 unpublished books. A total size is, is, 5 4.5Â GB was released in June of 2018.

103

00:13:01.140 --> 00:13:07.969

Didong Li: It was trained on 8, p. 600 gpus, which which was the state of the art. Back to that years

104

00:13:07.990 --> 00:13:15.539

Didong Li: I was trained for 30 days, you know it's it's pretty crazy, actually, if you, even if you look at these numbers nowadays, it's still crazy.

105

00:13:16.570 --> 00:13:24.740

Didong Li: And then, about 6 months or 8 months later, in 2019 February, we released the second version, Gpt. 2,

106

00:13:25.351 --> 00:13:32.089

Didong Li: which is the almost the same as Gp. One. The architecture, but is a modified normalization

107

00:13:33.227 --> 00:13:37.490

Didong Li: the number of parameters is about 10 times the size of Gbt. One.

108

00:13:37.520 --> 00:13:45.519

Didong Li: The training data size is 40Â GB is much bigger, and the training cost is about 10 times the cost for Gbt. One.

109

00:13:46.950 --> 00:13:54.870

Didong Li: This is not the end of the story, and then they further update it to Gp 3. And this is about the

110

00:13:55.240 --> 00:14:10.189

Didong Li: major changes in this table. Where they modify the architecture to make it more scalable. So the number of parameters is about a hundred times the parameters in Gpt. 2. It's or a hundred 1 billion.

111

00:14:10.450 --> 00:14:14.289

Didong Li: And the training sample size was 5 70Â GB.

112

00:14:14.560 --> 00:14:19.579

Didong Li: It's this will train our English Wikipedia. It's almost everything pretty big.

113

00:14:19.640 --> 00:14:28.759

Didong Li: Released in 2020 May of 2020. The training cost is again like 300 times the cost of Gpt. 2.

114

00:14:29.850 --> 00:14:42.020

Didong Li: Then there is an intermediate version called Gpt. 3.5. That's actually the Gpt we are more familiar with. When you say Gpt, you really refer to 3.5 or later.

115

00:14:42.690 --> 00:14:47.310

Didong Li: It was declosed in 2022 March.

116

00:14:47.450 --> 00:14:57.620

Didong Li: and the same number of parameters, which is surprising because our time number parameters is increased a lot, but this time number of parameters are same, but

117

00:14:57.630 --> 00:15:02.140

Didong Li: we don't even know the architecture. We don't know the training data or the training cost.

118

00:15:03.450 --> 00:15:28.020

Didong Li: Everything is so is a magic. So they must did something non trivial, architecture wise. That's why they can use the same number of parameters that made make the performance much, much better. And so why, it's not the close. My conjecture is this is the 1st like business product they use, which is the chat to be you tried at the beginning. So they wanna keep. Keep it secret even even for today.

119

00:15:28.150 --> 00:15:42.450

Didong Li: And you may wonder, wait a minute. This is this is not true, because back to 2022 March. I didn't see China Gpt at all. I 1st used that in November of 2022, maybe even later. So why, there is a gap.

120

00:15:42.460 --> 00:15:51.409

Didong Li: So this Gpt. 3.5 is the base model for Chat Gpt, but it's not exactly the same as Chat Gpt. I'll explain the difference on the next slide.

121

00:15:51.710 --> 00:15:53.290

Didong Li: But there's a time, Cap.

122

00:15:53.860 --> 00:15:58.059

Didong Li: and so last year 2023 in March.

123

00:15:58.220 --> 00:16:08.930

Didong Li: Gpt. 4, was released. This is the most updated model. If you open up your chat. Gpd, app if you use it today. It was built

124

00:16:08.980 --> 00:16:11.170

Didong Li: on this Gpt. 4 model.

125

00:16:11.260 --> 00:16:17.769

Didong Li: There are many different versions, but they are all under this Gp. 4. Architecture is also trained

126

00:16:18.020 --> 00:16:38.149

Didong Li: for text prediction, and also use the reinforcing learning human feedback. It accepts both text and images. If you recall this, if you use Gbt like in 2022, you can only input text. And now you can upload figures or screenshots, or even Pdfs, we will see some examples today

127

00:16:38.230 --> 00:16:41.250

Didong Li: where we do upload Pdf and figures.

128

00:16:42.020 --> 00:17:00.720

Didong Li: and so the number, the number of parameter is still a secret. It's estimated as 1.7 trillion, which is much bigger than 3.5. I don't even know how people estimate it. I've seen different estimates, one trillion, 2 trillion, 1.7. This is from Wikipedia. So let's say, this is a ground truth in some sense.

129

00:17:01.620 --> 00:17:08.109

Didong Li: Again, we don't know the training data. We don't know the cost. There's estimate again, I don't know how how to estimate it.

130

00:17:08.790 --> 00:17:20.589

Didong Li: So this is the most recent version, and there are a lot of recent news like from the CEO or CTO of Openai, who said, Okay, TP. 5 is on the way.

131

00:17:20.800 --> 00:17:25.530

Didong Li: and it has been on the way for a year, so we don't know when 2 p. 5 can be released.

132

00:17:26.829 --> 00:17:32.009

Didong Li: That's it. So this is a very brief history of Gpt.

133

00:17:32.060 --> 00:17:34.669

Didong Li: And so the next slide I will show you.

134

00:17:34.730 --> 00:17:37.900

Didong Li: Chatgbt is not the only locked up with model.

135

00:17:38.300 --> 00:17:42.739

Didong Li: Well, this is a very competitive game, so this figure

136

00:17:43.040 --> 00:17:44.880

Didong Li: summarize some

137

00:17:45.450 --> 00:17:50.750

Didong Li: existing large language models. The X coordinate is time.

138

00:17:50.960 --> 00:17:59.290

Didong Li: Year is from 3 20 number 2 to 2024 March 20.th That's how

139

00:17:59.410 --> 00:18:04.980

Didong Li: that's the last day they collected data. I just screenshot this from a website.

140

00:18:05.880 --> 00:18:11.060

Didong Li: and the the Y coordinate here is the number of parameters in bidding.

141

00:18:12.180 --> 00:18:17.949

Didong Li: and so the size of each circle is the number of parameters.

142

00:18:18.390 --> 00:18:28.050

Didong Li: and you can see different colors. These colors are some major players in the game. For example, Amazon has this Olympus and Astropic

143

00:18:28.616 --> 00:18:36.553

Didong Li: which is. You know, most employees are from Openai. They they have this new company.

144

00:18:37.150 --> 00:18:42.229

Didong Li: apple Chinese. There's not a single player. There are multiple players in China.

145

00:18:42.510 --> 00:18:43.710

Didong Li: Google.

146

00:18:43.920 --> 00:18:48.999

Didong Li: Meta, or Meta model is open source. That's very different from other models.

147

00:18:49.120 --> 00:18:51.359

Didong Li: Microsoft, Openai.

148

00:18:51.520 --> 00:18:52.690

Didong Li: and others.

149

00:18:52.720 --> 00:18:56.019

Didong Li: Apparently we can see some trends, for example.

150

00:18:56.310 --> 00:19:00.139

Didong Li: the models are getting bigger and bigger overall. You can see

151

00:19:00.380 --> 00:19:01.770

Didong Li: this monster.

152

00:19:02.480 --> 00:19:04.759

Didong Li: It has 2 training parameters.

153

00:19:06.760 --> 00:19:08.489

Didong Li: another trend is

154

00:19:08.580 --> 00:19:18.649

Didong Li: so if you go from left to right, even if the overall trend is, the model is getting bigger and bigger, there are still some research going on focusing on smaller models.

155

00:19:18.790 --> 00:19:26.429

Didong Li: So, for example, these type of models have about 1 billion parameter. This is something actually, we can use on our laptop.

156

00:19:27.420 --> 00:19:31.850

Didong Li: So there were some companies like huge companies focus on bigger

157

00:19:32.271 --> 00:19:44.190

Didong Li: bigger models like Meta, they have huge computing resources. But there are some companies or startups focusing on the smaller one, maybe specifically designed for some domain tasks.

158

00:19:45.591 --> 00:19:48.400

Didong Li: Then, on the top row.

159

00:19:48.590 --> 00:19:50.000

Didong Li: these are box.

160

00:19:50.630 --> 00:19:57.929

Didong Li: So what are both? No? So the the large large language models I showed you on the previous slide or the so-called base model.

161

00:19:57.990 --> 00:20:01.660

Didong Li: It was trained to predict the next sentence.

162

00:20:01.700 --> 00:20:03.170

Didong Li: it's like a

163

00:20:03.300 --> 00:20:06.510

Didong Li: educated person, but without any major.

164

00:20:06.860 --> 00:20:13.889

Didong Li: Then we would like to train the model further, train the model so that the model has some domain knowledge.

165

00:20:13.910 --> 00:20:18.149

Didong Li: For example, what is Chat gpt chat gpt is a

166

00:20:18.580 --> 00:20:21.260

Didong Li: a person major in conversation.

167

00:20:21.430 --> 00:20:28.159

Didong Li: So you further fine-tune your base model so that it's more good at chatting with people.

168

00:20:28.240 --> 00:20:43.890

Didong Li: and you can use other Gpts. I will show you some other examples later. For example, there's a coding version of Chat Gpt. Called Codex is Gpt. Plus coding, major. It's the Commercial science major of of a larger model. This is a conversation.

169

00:20:43.930 --> 00:20:45.510

Didong Li: major tragedy team.

170

00:20:46.210 --> 00:21:04.320

Didong Li: and so you may wonder where is Chat Gbt. Chat Gpt was released in November of 2022. It was a fine tuned model for conversational use, based on Gpt. 3.5. So Gpt. 3.5 was released in march of 2022. The Chat Bot

171

00:21:04.430 --> 00:21:06.590

Didong Li: was born a couple of months later.

172

00:21:06.970 --> 00:21:09.049

Didong Li: which which is here actually

173

00:21:13.020 --> 00:21:18.530

Didong Li: and about the same time, you can see a popularity in

174

00:21:18.710 --> 00:21:23.510

Didong Li: in this large language model, because Gb. Chat, Gbd. Was so successful

175

00:21:23.620 --> 00:21:47.099

Didong Li: as we have seen the previous surveys. You know it's it's very well known even to general audience. So before this there are some breakthroughs in in deep learning, for example, Alphafold, like Alphago, they are very surprising, but they are mainly for like academia. But this attribute is really understandable to general audience. It has a huge impact to human.

176

00:21:47.340 --> 00:21:48.830

Didong Li: no human being.

177

00:21:48.870 --> 00:21:58.199

Didong Li: And so when you want to train this large model, or you want to play with it. You need some gpus which is a key component in a model training.

178

00:21:58.430 --> 00:21:59.450

Didong Li: And

179

00:21:59.520 --> 00:22:04.900

Didong Li: so Nvidia is a leader, a leader company in this Gpu

180

00:22:05.890 --> 00:22:08.760

Didong Li: a Gpu market, and you can see

181

00:22:09.238 --> 00:22:14.979

Didong Li: by the time of Chat was was born there is a huge increase in the stock

182

00:22:15.050 --> 00:22:26.410

Didong Li: price is probably 10 times as the price at that at that time, and there's a famous joke by the CEO of Media, Jesse Hong. The more Gpu use you buy, the more money you save.

183

00:22:26.590 --> 00:22:30.890

Didong Li: It's it's like a joke. But sometimes, you know, in some sense it's true.

184

00:22:31.730 --> 00:22:35.370

Didong Li: I'm I'm not advertising for Nvidia, by the way. Well.

185

00:22:35.510 --> 00:22:36.420

Didong Li: oh.

186

00:22:36.830 --> 00:22:38.518

Didong Li: so okay, that's a

187

00:22:39.580 --> 00:22:54.489

Didong Li: That's a brief history of of Chat Gpt, and hopefully I explain the the difference between Gpt and Chat Gpt, and if you do not like chat, gpt, there are many different chat box.

188

00:22:54.840 --> 00:23:04.649

Didong Li: Other Chatbot, if you want to try them. So on this last column you can see some to be announced models they are on the way.

189

00:23:04.670 --> 00:23:11.739

Didong Li: But because this this figure was made in march. And actually Lama 3 was released about like 2 weeks ago.

190

00:23:11.870 --> 00:23:14.629

Didong Li: So it's already there, as much bigger as has

191

00:23:15.070 --> 00:23:18.099

Didong Li: of 405 billion parameters

192

00:23:18.240 --> 00:23:21.939

Didong Li: and 25, unfortunately, is still on the way.

193

00:23:23.530 --> 00:23:26.969

Didong Li: Okay? So looking back to Chat Gpt.

194

00:23:27.840 --> 00:23:34.640

Didong Li: I do have this figure. It's a screenshot from the 1st orchard of Ktbd, which was based on 3.5.

195

00:23:34.830 --> 00:23:43.720

Didong Li: It claims that this is this is claimed by Openai themselves. No capabilities of Chat Gbt. There are so many, and I couldn't find such a

196

00:23:43.950 --> 00:23:49.299

Didong Li: a summary. 1484. 0, but it should be the least would be longer.

197

00:23:49.590 --> 00:23:56.510

Didong Li: It can do a lot of things but related to our writing workshop. I I highlighted some

198

00:23:56.770 --> 00:24:02.189

Didong Li: capabilities. For example, it can do grammar correction. We will try this today

199

00:24:03.190 --> 00:24:07.420

Didong Li: or explain code again. We will try this last section.

200

00:24:08.730 --> 00:24:35.859

Didong Li: It can track keywords or summarize a paper. You can summarize the paper in a paragraph or in my sentence like this. Tldr, or you can just extract some keywords. Sometimes it can generate title for for your papers, for example. Myself, I'm not really good at making titles for paper. Sometimes I write a paper first, st and they ask Gpt to generate or suggest some titles for me. Sometimes the titles are pretty pretty funny.

201

00:24:37.138 --> 00:24:46.419

Didong Li: It can be used to create tables from tax, or it can generate a code for tables or figures.

202

00:24:47.850 --> 00:25:09.280

Didong Li: it can also translate programming languages, you know. Again myself as an example. I was trained as a pure mathematician at the beginning, so I I use Mathematica at that time. I don't know any r at all. Then, during my grad school, I slowly changed to apply math, where I use my lab every day.

203

00:25:09.640 --> 00:25:31.910

Didong Li: but I end up being in a Bellstead department as as a faculty. It's it's pretty challenging to me. So everyone use R in my department. So I have to use R sometimes in the teaching I need to use. R. So I need to learn a new language. And you know, in Bellsta Department students are required. At least I. Unc. Students are required to learn sauce cause Sars is very, very popular in pharmaceutical companies or in in

204

00:25:32.010 --> 00:25:34.220

Didong Li: in government, for example, FDA.

205

00:25:34.230 --> 00:25:47.948

Didong Li: So there are so many languages going on. They have different pros and cons, and sometimes one algorithm, you can find the package only in one language, now the other. So sometimes you would like to translate this.

206

00:25:48.420 --> 00:25:55.170

Didong Li: this code between different languages. So we will try some of these today in this presentation.

207

00:25:56.090 --> 00:25:58.119

Didong Li: And so let's go.

208

00:25:58.410 --> 00:26:11.510

Didong Li: Could you please scan this QR code, I created a Google Doc, because when we play with some sample examples together, we better use the same tags. So everything is in the Google, Doc.

209

00:26:14.230 --> 00:26:18.750

Didong Li: let me know if you do not have access, and I can change the setup.

210

00:26:25.000 --> 00:26:29.299

Didong Li: Apologize. We're trying to get it fixed.

211

00:26:30.110 --> 00:26:30.990

Didong Li: Sorry.

212

00:26:38.720 --> 00:26:39.630

Didong Li: or

213

00:26:45.440 --> 00:26:46.650

Didong Li: I see.

214

00:26:52.610 --> 00:26:57.539

Didong Li: I guess you need the app to edit it. But if you want to read it, you you need the app.

215

00:26:58.340 --> 00:26:59.380

Didong Li: I suppose.

216

00:27:02.300 --> 00:27:03.160

Didong Li: little bit

217

00:27:04.750 --> 00:27:08.029

Didong Li: so ideally, you can open this Google Doc on your laptop.

218

00:27:08.210 --> 00:27:11.280

Didong Li: Or if you want to use your cell phone is is fine.

219

00:27:11.920 --> 00:27:14.489

Didong Li: If you prefer to use Cbd on your phone

220

00:27:18.390 --> 00:27:19.690

Didong Li: 45, 50.

221

00:27:20.980 --> 00:27:24.959

Didong Li: Oh, I don't think it's that that long

222

00:27:25.660 --> 00:27:26.460

Didong Li: oops.

223

00:27:28.270 --> 00:27:30.300

Didong Li: This is the

224

00:27:30.880 --> 00:27:33.480

Didong Li: the 9 pages long.

225

00:27:49.160 --> 00:27:49.960

Didong Li: Looks

226

00:27:51.160 --> 00:27:53.070

Didong Li: okay. The same with you as you.

227

00:27:53.240 --> 00:27:54.260

Didong Li: Okay, cool

228

00:27:56.509 --> 00:28:02.440

Didong Li: because later on we need to copy something to your chat. Gpt, and we will do this together.

229

00:29:40.220 --> 00:29:41.260

Didong Li: Yeah, the balls is.

230

00:29:43.080 --> 00:29:43.750

Didong Li: oh.

231

00:29:44.260 --> 00:29:45.929

Didong Li: that's less remote.

232

00:29:48.340 --> 00:29:49.910

Didong Li: you know, access to the

233

00:29:50.520 --> 00:29:52.749

Didong Li: Google, Doc. Alright, great.

234

00:29:52.910 --> 00:29:56.959

Didong Li: Okay, let's move on. So the 1st task you wanna try

235

00:29:57.020 --> 00:29:59.560

Didong Li: is to summarize a research paper.

236

00:30:00.395 --> 00:30:02.669

Didong Li: You know. Sometimes you have.

237

00:30:02.930 --> 00:30:25.870

Didong Li: You find a lot of papers of interest, and you don't have time to read off them. For example, if you open your open up your Google scholar. Sometimes there are some recommended papers, and you wanna choose some of them to read in details, and you may download them and ask Gpt to give you a summary. It's it's usually more than the abstract, and so that you can decide which one to read.

238

00:30:26.740 --> 00:30:27.770

Didong Li: Oh.

239

00:30:28.260 --> 00:30:33.839

Didong Li: so I was. I was thinking about like what paper to use in this in this discussion

240

00:30:33.980 --> 00:30:46.909

Didong Li: at the beginning. I thought, maybe we can use a paper by myself, but that may be embarrassing. So I decided to look at the paper written by smarter people. So I opened up the website of Jasa, a statistical journal.

241

00:30:47.200 --> 00:30:48.070

Didong Li: I'll

242

00:30:48.260 --> 00:31:07.890

Didong Li: and I found that they released some recently accepted papers written by very smart people. So this one was accepted or posted online in July 24. So it's a very recent paper. So I said, Okay, I thought about this, why not? You just use use this paper and to ask Gpt to write a summary.

243

00:31:08.700 --> 00:31:15.619

Didong Li: And so the title here is matrix completion, we mean is not at random, and it's applications in causal panel data models.

244

00:31:15.960 --> 00:31:36.079

Didong Li: And, to be honest, I have no idea what does mean. I'm not expert in matrix completion or missing data or causal panel data model. I have no idea if you happen to know this area perfect, then you can check whether the summary is good or not. But if you are not expert in this field, that's totally okay, I don't understand anything about this paper at all. So

245

00:31:37.226 --> 00:31:53.859

Didong Li: so the because I don't think Jasa is open access, and you need to download it. And that's why I have a a Google Google drive link in the Google Doc. So if you click on the link, you can download the paper, I download it from from Unc website.

246

00:31:53.960 --> 00:31:57.900

Didong Li: So this is not a business use. I guess it's okay, but correct.

247

00:31:58.250 --> 00:31:59.380

Didong Li: Oh, okay.

248

00:31:59.750 --> 00:32:03.599

Didong Li: So please download this paper through the Google drive.

249

00:32:03.640 --> 00:32:05.900

Didong Li: Let me know if you cannot download it.

250

00:32:06.670 --> 00:32:08.390

Didong Li: As a Pdf is

251

00:32:08.800 --> 00:32:10.490

Didong Li: 45 pages. Is that correct?

252

00:32:31.170 --> 00:32:32.540

Didong Li: So

253

00:32:36.740 --> 00:32:38.529

Didong Li: I don't know which one is. The

254

00:32:41.676 --> 00:32:42.483

Didong Li: okay?

255

00:32:55.570 --> 00:32:56.240

Didong Li: Yeah.

256

00:33:02.570 --> 00:33:04.070

Didong Li: If it's just me.

257

00:33:07.270 --> 00:33:08.509

Didong Li: did you download it?

258

00:33:09.880 --> 00:33:10.660

Didong Li: Yeah.

259

00:33:12.900 --> 00:33:15.080

Didong Li: Why, I'm stuck here.

260

00:33:17.620 --> 00:33:18.670

Didong Li: Let's go to

261

00:33:18.770 --> 00:33:20.350

Didong Li: Chatebt Whatsapp.

262

00:33:21.800 --> 00:33:26.470

Didong Li: If you do not have account, let let me know. Otherwise we can.

263

00:33:27.750 --> 00:33:29.159

Didong Li: We do not have a

264

00:33:29.990 --> 00:33:32.120

Didong Li: can you register. Now it's free.

265

00:33:33.480 --> 00:33:34.300

Didong Li: Thanks.

266

00:33:36.830 --> 00:33:39.040

Didong Li: It's called Chat gpt.com

267

00:33:39.400 --> 00:33:40.299

Didong Li: loves that.

268

00:33:44.690 --> 00:33:46.429

Didong Li: Okay, now let me think.

269

00:33:48.180 --> 00:33:51.479

Didong Li: If you do have account, you click this icon.

270

00:33:52.280 --> 00:33:55.069

Didong Li: This allows you to upload a file

271

00:33:55.430 --> 00:33:59.550

Didong Li: they support like figures. Pdfs, the one on the left.

272

00:34:00.340 --> 00:34:01.840

Didong Li: It's a misattachment

273

00:34:02.000 --> 00:34:04.520

Didong Li: and send the paper

274

00:34:14.360 --> 00:34:15.660

Didong Li: that's not there.

275

00:34:20.610 --> 00:34:21.400

Didong Li: Out

276

00:34:25.429 --> 00:34:26.230

Didong Li: oops

277

00:34:39.010 --> 00:34:40.689

Didong Li: you download the paper first, st

278

00:34:41.030 --> 00:34:42.910

Didong Li: and then you go to your

279

00:34:45.570 --> 00:34:46.360

Didong Li: oops.

280

00:34:46.530 --> 00:34:48.140

Didong Li: Go to church activity.

281

00:34:48.409 --> 00:34:50.439

Didong Li: And you do you see this icon

282

00:34:51.100 --> 00:34:52.080

Didong Li: attachment?

283

00:34:56.949 --> 00:34:57.740

Didong Li: Let me.

284

00:35:04.690 --> 00:35:06.300

Didong Li: Okay, that's it.

285

00:35:12.420 --> 00:35:13.420

Didong Li: You ask

286

00:35:13.530 --> 00:35:16.929

Didong Li: whatever question you like. For example, summarize

287

00:35:16.970 --> 00:35:19.010

Didong Li: this paper for this.

288

00:35:21.060 --> 00:35:22.839

Didong Li: you can use your own form.

289

00:35:32.650 --> 00:35:33.680

Didong Li: Let's go.

290

00:35:39.535 --> 00:35:43.540

Didong Li: Yeah, it's a lot. Let's let's let's discuss this.

291

00:35:58.590 --> 00:36:01.090

Didong Li: We already have a comment. It's a lot

292

00:36:01.270 --> 00:36:03.030

Didong Li: Gpt is

293

00:36:03.530 --> 00:36:11.649

Didong Li: really worrying sometimes. Give us, repeated repetitive quest answers. You know, it's it's it's it's not really condensed.

294

00:36:12.620 --> 00:36:13.380

Didong Li: They

295

00:36:14.680 --> 00:36:21.820

Didong Li: I don't think so. You know. You can see mine is the same as yours.

296

00:36:22.960 --> 00:36:24.169

Didong Li: Not at all right.

297

00:36:24.670 --> 00:36:25.530

Didong Li: This is

298

00:36:26.570 --> 00:36:30.580

Didong Li: summary by section. In my case it gives a summary of the key takeaways.

299

00:36:30.850 --> 00:36:36.149

Didong Li: my multi-sexualized, but it's very different. And basically it was the same concept

300

00:36:36.700 --> 00:36:38.150

Didong Li: last 15.

301

00:36:39.430 --> 00:36:42.099

Didong Li: Yes, I did this on on purpose.

302

00:36:44.100 --> 00:36:45.000

Didong Li: So

303

00:36:45.330 --> 00:36:51.690

Didong Li: so before before we discuss this, I mean, let me share with you a story. Thanks. This is a good point.

304

00:36:51.750 --> 00:36:53.149

Didong Li: So we

305

00:36:53.430 --> 00:36:57.111

Didong Li: we were doing a research project early this year.

306

00:36:58.260 --> 00:37:02.170

Didong Li: that was from a national survey on hypertension.

307

00:37:02.520 --> 00:37:10.010

Didong Li: So we we would like to predict whether this person is at a high risk of hypertension or not

308

00:37:10.030 --> 00:37:13.129

Didong Li: due to their child's related features.

309

00:37:14.050 --> 00:37:16.590

Didong Li: and 2 of my Phd. Students were.

310

00:37:17.559 --> 00:37:21.035

Didong Li: try to use large language models to predict

311

00:37:21.590 --> 00:37:33.849

Didong Li: high risk or not. It's essentially a binary classification problem. You, input the house features output is yes or no, yes means at risk of hypertension. No means, not at at risk.

312

00:37:34.610 --> 00:37:44.320

Didong Li: And so what we did was we we use a base model called Q win. It's a larger model developed by Alibaba with

313

00:37:44.390 --> 00:38:12.180

Didong Li: 7 billion parameters, 3 billion. I remember that 2 versions. And then we fine tune this model based on some sample questions. Say, I, input the the house feature of about a hundred people. I tell Gpt to answer, Okay, yes, for this one, no, for this one. It's like A, Q&A, we use this 100 training samples to fine tune the model. And then we, we use 50,000 as the test data to check the class classification accuracy.

314

00:38:12.180 --> 00:38:22.860

Didong Li: And in the training data you have to design questions. Right? Our question was, does this paper does this person has a high risk of hypertension. Here's the person's house feature.

315

00:38:24.420 --> 00:38:26.970

Didong Li: and the performance was really bad.

316

00:38:27.260 --> 00:38:35.170

Didong Li: The accuracy was like 50%, which is random gas. Then my students added this period

317

00:38:36.240 --> 00:38:38.440

Didong Li: to the training in question.

318

00:38:38.710 --> 00:38:40.929

Didong Li: the accuracy was like 75%.

319

00:38:41.880 --> 00:38:42.980

Didong Li: So it's

320

00:38:43.040 --> 00:39:02.140

Didong Li: you know you, you can tell. You know, Charity Pt is not really to understand the mathematical model behind this. But it's really a language model. It's I translate everything here into like tokens into language, and then do some like a tax embedding. I run something in a black box.

321

00:39:02.340 --> 00:39:08.250

Didong Li: So even if you change the the period with something like you have very different different result. That's

322

00:39:08.590 --> 00:39:14.646

Didong Li: so real clearance. I didn't make it up. Actually, I'm happy to send you the the report. We have technical report of that.

323

00:39:15.290 --> 00:39:29.905

Didong Li: okay. So so, looking back into this, you know, we have seen multiple formats. For example, some some of us get this summary section by section. Some of us have these key takeaways. Is that correct?

324

00:39:30.310 --> 00:39:43.130

Didong Li: I I guess so. Personally, I I don't think it matters like what was the what the format is. So the the important thing here is whether is is accurate was correct. At least

325

00:39:43.280 --> 00:39:46.529

Didong Li: we want to make sure it doesn't generate something very crazy.

326

00:39:46.700 --> 00:39:47.670

Didong Li: So

327

00:39:48.830 --> 00:40:00.039

Didong Li: so how about we read this? Read your own answer, and let me know whether you think it makes sense or not, that it sounds like a just a paper as a research paper or not.

328

00:40:00.870 --> 00:40:02.130

Didong Li: I'll read mine.

329

00:40:08.240 --> 00:40:13.659

Didong Li: So basically chat, Gpt, what you are saying is just a language to

330

00:40:13.690 --> 00:40:15.630

Didong Li: it's not a lighter

331

00:40:16.150 --> 00:40:25.409

Didong Li: analytical to it can analyze the data, but it analyze the data through language models.

332

00:40:26.510 --> 00:40:36.940

Didong Li: Like, I have all the data like demographics. And like, in case I have hesitation. I have medication.

333

00:40:37.590 --> 00:40:46.719

Didong Li: I know, and the age. So I want to say, I have tragedy. Can you analyze the age?

334

00:40:46.910 --> 00:41:01.820

Didong Li: Yeah, so will you give you the I'll say you give me the correct answer is that probability? But it's not doing that analysis as you expected. For example, if you ask what is one plus one?

335

00:41:01.900 --> 00:41:04.319

Didong Li: You will get 2. This high probability.

336

00:41:04.870 --> 00:41:09.320

Didong Li: it's not really doing one plus one. It's translate one plus one

337

00:41:10.060 --> 00:41:15.789

Didong Li: into the tokens are treated as a language and invited to a very high dimensional vector space.

338

00:41:16.170 --> 00:41:17.610

Didong Li: and then do some

339

00:41:17.810 --> 00:41:20.390

Didong Li: prediction there. So what's the

340

00:41:20.730 --> 00:41:23.509

Didong Li: next word? Is the highest probability.

341

00:41:23.820 --> 00:41:24.820

Didong Li: it's true.

342

00:41:25.120 --> 00:41:26.789

Didong Li: And then 8 output 2.

343

00:41:27.740 --> 00:41:34.740

Didong Li: It kept trying to use it to analyze my data. Yeah. So for that of analysis, I think it, it works.

344

00:41:35.020 --> 00:41:36.479

Didong Li: But it's not as

345

00:41:36.850 --> 00:41:40.940

Didong Li: the the the steps are not as as human being.

346

00:41:43.000 --> 00:41:49.490

Didong Li: So this is a little bit beyond the scope of this. I'm happy to chat offline. It's pretty complicated how it works.

347

00:41:50.530 --> 00:41:53.810

Didong Li: Does this replace the statistician?

348

00:41:54.398 --> 00:42:03.709

Didong Li: Well, there should be some panel discussion I just have on this. You're you're welcome to join. There's like a statistical AI panel. Discussion.

349

00:42:05.210 --> 00:42:07.699

Didong Li: it might be on, it might be on Monday. Right

350

00:42:07.900 --> 00:42:09.479

Didong Li: should be a panel discussion.

351

00:42:09.890 --> 00:42:14.950

Didong Li: But this is like a writing workshop, so I won't go too deep into the technical side.

352

00:42:15.370 --> 00:42:16.040

Didong Li: Ben.

353

00:42:16.150 --> 00:42:26.960

Didong Li: What what does it do for the like? The actual latex in here so like when there's these equations. Is it converting those into text, or is it skipping over those? Or I think it's good?

354

00:42:27.160 --> 00:42:32.290

Didong Li: It skips over? Yes, yes, yes.

355

00:42:34.980 --> 00:42:41.280

Didong Li: yes. So each time there's a randomness in awkward. So each time they, of course, we.

356

00:42:41.390 --> 00:42:47.660

Didong Li: we will get different results. So how to like, compare or combine those 3. Also they are not very much.

357

00:42:48.500 --> 00:42:55.640

Didong Li: Oh, that's a that's a that's a good question. I think it's it's case by case, you know, for this type of summary. I guess

358

00:42:56.000 --> 00:43:02.329

Didong Li: if you if you have, say 3 different summaries, probably it doesn't really matter which one which I use.

359

00:43:02.988 --> 00:43:05.280

Didong Li: But say, if you. Wanna

360

00:43:05.390 --> 00:43:10.629

Didong Li: if you want to predict hypertension and you try this 3 times, you have different results. That that's tricky.

361

00:43:11.130 --> 00:43:16.680

Didong Li: I I don't really have a clear answer to this. You know, it's it's not supposed it's a chat model.

362

00:43:16.880 --> 00:43:26.319

Didong Li: If you want, you won't use it to for something else. For data analysis, you, I personally suggest you to use some other fine tune model for different tasks.

363

00:43:26.700 --> 00:43:27.906

Didong Li: for example.

364

00:43:28.540 --> 00:43:35.469

Didong Li: you know there's a medical Gpt. You can predict disease or not is is fine tuned for that task.

365

00:43:35.750 --> 00:43:39.309

Didong Li: We can't use Chat Gp. To do everything. It's just for chat.

366

00:43:39.440 --> 00:43:41.309

Didong Li: I would say. It's more good at chatting.

367

00:43:42.330 --> 00:43:43.070

Didong Li: Thanks.

368

00:43:46.120 --> 00:43:47.380

Didong Li: Finish reading.

369

00:43:48.550 --> 00:43:50.050

Didong Li: So who they?

370

00:43:50.250 --> 00:43:53.649

Didong Li: If your summary doesn't make sense.

371

00:43:56.930 --> 00:43:57.870

Didong Li: Okay.

372

00:43:58.400 --> 00:43:59.460

Didong Li: that means

373

00:43:59.610 --> 00:44:11.612

Didong Li: it's okay. It's it's a reasonable summary but you know, sometimes we have some complaints like, why so long can we make it shorter? Maybe we can try

374

00:44:13.110 --> 00:44:15.890

Didong Li: when you make a sorter

375

00:44:16.020 --> 00:44:17.065

Didong Li: summarize?

376

00:44:18.380 --> 00:44:19.740

Didong Li: Oh.

377

00:44:29.410 --> 00:44:33.430

Didong Li: so essentially, it's like you are training with. With a kid.

378

00:44:34.808 --> 00:44:41.580

Didong Li: You just tell Gpt what you want, what you like, what you dislike, and then you communicate.

379

00:44:42.100 --> 00:44:50.319

Didong Li: Keep, keep communicating you have, you may have a a result that better meets your expectation. So this time is is shorter.

380

00:44:51.040 --> 00:44:59.889

Didong Li: Those 3 points. One is about methodology, application, historical results. And actually, that's probably the 3 most important components. You know, data.

381

00:45:00.120 --> 00:45:03.050

Didong Li: It's it's called serial master. And you already have application.

382

00:45:03.450 --> 00:45:06.409

Didong Li: So like, what what type of setting. Would you

383

00:45:07.630 --> 00:45:12.100

Didong Li: use this? I like, you're not getting

384

00:45:12.230 --> 00:45:16.369

Didong Li: the important technical, administrative summary like I, I

385

00:45:16.470 --> 00:45:20.630

Didong Li: feel like this is helpful information. If you're looking for a bunch of articles to cite

386

00:45:20.850 --> 00:45:24.120

Didong Li: on a general topic in an introduction to

387

00:45:24.330 --> 00:45:30.519

Didong Li: a somewhat relevant paper. But like, is that the best case for this, or

388

00:45:30.680 --> 00:45:34.900

Didong Li: because you're not really getting information about what it's about.

389

00:45:36.630 --> 00:45:41.959

Didong Li: Yeah, that that's a good question. So so one user case I can think of is, you know, I

390

00:45:42.350 --> 00:45:44.299

Didong Li: 10 papers on my desk.

391

00:45:44.370 --> 00:45:45.950

Didong Li: I'm underneath my office.

392

00:45:46.280 --> 00:45:47.240

Didong Li: and

393

00:45:47.990 --> 00:46:00.599

Didong Li: I can also ask if you generate all these summaries. I can read it and decide which one that's that's 1 simple case warning. Sometimes you ask for students before faculty to summarize a paper.

394

00:46:01.490 --> 00:46:06.399

Didong Li: You want to read the entire paper, look at 5 days and check whether your students aren't make sense.

395

00:46:07.690 --> 00:46:10.429

Didong Li: I know sometimes people use this to write

396

00:46:10.850 --> 00:46:14.659

Didong Li: correct paper reviews. If you are review of a paper.

397

00:46:14.820 --> 00:46:19.550

Didong Li: and you already at the beginning, you summarize the paper. I know someone used this, you know, this is

398

00:46:20.810 --> 00:46:22.400

Didong Li: or not advertising

399

00:46:23.020 --> 00:46:29.660

Didong Li: this. This actually used to be reviewed. Paper. There are actually some policies, some businesses that have seen these type of use.

400

00:46:33.270 --> 00:46:35.990

Didong Li: Okay, I think this one is a fairly

401

00:46:36.400 --> 00:46:39.960

Didong Li: good summary, at least my point of view.

402

00:46:41.110 --> 00:46:46.750

Didong Li: you know. Sometimes you can. You can either. change, change your problem. Say.

403

00:46:46.810 --> 00:46:56.369

Didong Li: can you give me the 3 key takeaways of this paper or 2 takeaways. You can just make your question very specific if you give you maybe

404

00:46:56.400 --> 00:46:58.859

Didong Li: a more ideal answer.

405

00:47:00.370 --> 00:47:01.580

Didong Li: Oh.

406

00:47:01.750 --> 00:47:04.350

Didong Li: okay, let's move on to the next.

407

00:47:05.340 --> 00:47:08.740

Didong Li: I'll skip some. I'll prepare something in case it's

408

00:47:08.930 --> 00:47:10.800

Didong Li: it's not working at all.

409

00:47:11.220 --> 00:47:21.270

Didong Li: Okay? So the next one is to write an outline for manuscript. I'll answer the question before you. We go into the the discussion.

410

00:47:22.060 --> 00:47:25.980

Didong Li: This might be helpful if you are writing your first, st ever.

411

00:47:26.280 --> 00:47:29.909

Didong Li: so you have no idea. What does a paper look like

412

00:47:30.090 --> 00:47:38.309

Didong Li: like? What are the sections in the paper? What's the order of the sections? Or in case you miss any important sections.

413

00:47:38.450 --> 00:47:46.730

Didong Li: and maybe you can ask Gpt to write a very brief outline. It is. It's not detailed at all, but it give you like a a framework.

414

00:47:47.030 --> 00:48:01.090

Didong Li: and can improve the efficiency. So when I work with my students, if I'll ask this your 1st time writing the paper, if the answer is, yes, I would talk about the outline by myself in the in the meeting. Okay.

415

00:48:01.140 --> 00:48:07.029

Didong Li: we can start from this section. The next section. We talk about each of each information, each section.

416

00:48:07.080 --> 00:48:11.880

Didong Li: But now the student can simply ask Gpt the same question. So they don't have to ask me.

417

00:48:11.920 --> 00:48:13.050

Didong Li: And so

418

00:48:13.280 --> 00:48:31.230

Didong Li: so let's let's use the same example as as people say. Imagine we are the authors of this just paper, and we, before we write the paper, we want to ask Gpt to write a outline. For example, we decided the title.

419

00:48:31.530 --> 00:48:35.879

Didong Li: Then my question is, okay, can you write an outline for a paper titled this.

420

00:48:36.320 --> 00:48:43.720

Didong Li: So here's a question. Can you write an outline for the paper titled this?

421

00:48:44.090 --> 00:48:46.010

Didong Li: Let's let's try this here

422

00:48:47.730 --> 00:48:48.870

Didong Li: together.

423

00:48:50.080 --> 00:48:51.319

Didong Li: 1 14,

424

00:49:02.145 --> 00:49:04.339

Didong Li: you mean. Write the entire paper.

425

00:49:14.420 --> 00:49:15.550

Didong Li: Oh.

426

00:49:15.640 --> 00:49:16.840

Didong Li: well.

427

00:49:17.580 --> 00:49:26.350

Didong Li: I I was trying as a reference. But let's let's see the answer. Oh, can you write an outline

428

00:49:26.370 --> 00:49:30.250

Didong Li: of paper? I told

429

00:49:33.930 --> 00:49:35.090

Didong Li: question Mark.

430

00:49:46.010 --> 00:49:47.187

Didong Li: It's so long

431

00:49:59.060 --> 00:50:02.520

Didong Li: you can try this on your laptop. Or we can read this together.

432

00:50:02.580 --> 00:50:03.990

Didong Li: Okay, so

433

00:50:04.950 --> 00:50:06.120

Didong Li: abstract

434

00:50:08.220 --> 00:50:11.310

Didong Li: all real problems missing out. And random data.

435

00:50:12.320 --> 00:50:13.470

Didong Li: Sims

436

00:50:14.030 --> 00:50:16.430

Didong Li: structure seems. Okay. Intro.

437

00:50:16.710 --> 00:50:18.030

Didong Li: let it work

438

00:50:18.140 --> 00:50:19.680

Didong Li: mastology.

439

00:50:20.190 --> 00:50:21.900

Didong Li: theoretical results.

440

00:50:22.060 --> 00:50:23.280

Didong Li: application.

441

00:50:24.180 --> 00:50:26.609

Didong Li: discussion, conclusion. Usually it's a

442

00:50:26.670 --> 00:50:28.960

Didong Li: we merge this into one section.

443

00:50:29.370 --> 00:50:32.230

Didong Li: acknowledgment reference, appendix

444

00:50:32.720 --> 00:50:39.290

Didong Li: supplementary material including proofs, additional simulations, extended data analysis. So this looks like.

445

00:50:41.380 --> 00:50:44.419

Didong Li: real automa also is is pretty good.

446

00:50:44.440 --> 00:50:48.150

Didong Li: If you are not sure how to write a paper at all.

447

00:50:49.100 --> 00:51:17.799

Didong Li: you know, because when I when I was asked to present this, I don't know the target audience, you know. So, junior researchers, I'm a junior researcher or undergrad student is a junior researcher. So I I'm not sure who is the target audience. Maybe this is pretty redundant to you. Most of you know how to write a paper, but also it might be interest. It might be interesting if you, if you are advising a a undergrad students, for example, when they 1st read their their 1st paper. This might be, might be helpful

448

00:51:19.067 --> 00:51:20.710

Didong Li: any questions. Here

449

00:51:23.470 --> 00:51:26.910

Didong Li: is this outlined only based on the type of permission.

450

00:51:27.750 --> 00:51:34.750

Didong Li: Oh, not only so, it's based on its external information. It was trained on the entire

451

00:51:35.100 --> 00:51:44.169

Didong Li: Internet, in brief, almost all research papers, and it will tailor it for your title. That's why you can see a

452

00:51:45.090 --> 00:51:47.500

Didong Li: section 5 application.

453

00:51:48.160 --> 00:51:50.829

Didong Li: This is exactly the one in the in the title.

454

00:51:50.880 --> 00:51:52.690

Didong Li: And actually, that's a good point.

455

00:51:54.760 --> 00:51:59.549

Didong Li: If you want to make it more concrete, you can say, Okay, I will submit this paper to Jasa.

456

00:52:00.000 --> 00:52:01.330

Didong Li: Syrian method.

457

00:52:01.430 --> 00:52:09.299

Didong Li: Can you give me outline? The the result might be different, you know. Okay, I I you can tell. Okay, I'm I'm submitting this these 2 paper, this paper to new reps.

458

00:52:09.380 --> 00:52:16.559

Didong Li: The outline might be different, because different journals or different field have different style. Let's let's try this. Actually, that's a good point.

459

00:52:16.700 --> 00:52:17.290

Didong Li: So

460

00:52:19.360 --> 00:52:20.570

Didong Li: I,

461

00:52:20.620 --> 00:52:23.539

Didong Li: when to submit this paper

462

00:52:26.160 --> 00:52:29.890

Didong Li: to just theory and method.

463

00:52:30.130 --> 00:52:33.059

Didong Li: And you realize you're all trying.

464

00:52:43.150 --> 00:52:45.729

Didong Li: Okay, apparently knows when it's jazza.

465

00:52:46.260 --> 00:52:47.493

Didong Li: That's correct.

466

00:52:50.230 --> 00:52:51.599

Didong Li: I can't see

467

00:52:51.610 --> 00:52:53.030

Didong Li: big difference.

468

00:52:57.340 --> 00:52:58.689

Didong Li: Maybe the same.

469

00:53:05.390 --> 00:53:06.719

Didong Li: all this seems

470

00:53:07.240 --> 00:53:08.230

Didong Li: crat

471

00:53:08.770 --> 00:53:11.050

Didong Li: from the Jazza website, you know.

472

00:53:11.980 --> 00:53:16.139

Didong Li: Methodological rigor, theoretical contributions, practical applications.

473

00:53:16.260 --> 00:53:19.770

Didong Li: Okay? Oh, I've changed my mind.

474

00:53:20.230 --> 00:53:21.300

Didong Li: I

475

00:53:21.580 --> 00:53:24.450

Didong Li: I'd like to lose some meat on your ribs

476

00:53:27.230 --> 00:53:28.660

Didong Li: otherwise

477

00:53:29.220 --> 00:53:30.290

Didong Li: all time.

478

00:53:31.620 --> 00:53:35.046

Didong Li: maybe it's the same. No, if you say that means it doesn't work.

479

00:53:35.800 --> 00:53:36.910

Didong Li: let's see.

480

00:53:39.490 --> 00:53:41.042

Didong Li: Seems the same.

481

00:53:42.360 --> 00:53:43.590

Didong Li: Oh.

482

00:53:55.640 --> 00:53:56.680

Didong Li: well.

483

00:53:56.740 --> 00:54:00.280

Didong Li: there's a difference. Now there's an experiment session

484

00:54:01.220 --> 00:54:02.920

Didong Li: which was missing in the

485

00:54:03.450 --> 00:54:05.070

Didong Li: in the jazza outline.

486

00:54:05.170 --> 00:54:09.589

Didong Li: This is probably because the Gpt believes

487

00:54:10.586 --> 00:54:14.620

Didong Li: new reps audience focus more on experiment.

488

00:54:15.180 --> 00:54:16.710

Didong Li: No, yeah. Prep

489

00:54:17.220 --> 00:54:22.999

Didong Li: 2 applications in machine learning and data science. It can understand something about the context. But

490

00:54:23.060 --> 00:54:28.689

Didong Li: no, it's it's not perfect. Apparently it's it's not perfect, you know. The experiments overlap is

491

00:54:28.710 --> 00:54:31.980

Didong Li: with this application, especially these real world data.

492

00:54:32.370 --> 00:54:37.559

Didong Li: So probably you need to further talk to the model a little bit to make it better.

493

00:54:38.170 --> 00:54:39.450

Didong Li: Any other questions

494

00:54:43.490 --> 00:54:45.180

Didong Li: fantastic?

495

00:54:45.530 --> 00:54:51.860

Didong Li: Oh, oh, it's it, will you know I suggest you to

496

00:54:52.000 --> 00:54:55.229

Didong Li: upload the Jasa type.

497

00:54:57.164 --> 00:55:02.150

Didong Li: Oh, okay. So that means if you download the

498

00:55:02.400 --> 00:55:03.520

Didong Li: the South island.

499

00:55:04.613 --> 00:55:09.979

Didong Li: the videos and the figures.

500

00:55:10.410 --> 00:55:13.130

Didong Li: it's not bad. It works on worse. Yes.

501

00:55:13.370 --> 00:55:19.019

Didong Li: I think it depends on the training data. If you ask for a very popular journal company.

502

00:55:19.060 --> 00:55:24.130

Didong Li: as we ask the journal.

503

00:55:25.370 --> 00:55:27.826

Didong Li: yeah, thanks. That's great.

504

00:55:28.710 --> 00:55:30.860

Didong Li: beat on the back of the ph1. 0, cool.

505

00:55:31.952 --> 00:55:34.119

Didong Li: Let me share.

506

00:55:35.310 --> 00:55:36.625

Didong Li: And

507

00:55:38.950 --> 00:55:40.139

Didong Li: oh, it's recording.

508

00:55:40.430 --> 00:55:42.400

Didong Li: Still, it's still recording. Okay, cool.

509

00:55:43.450 --> 00:55:46.480

Didong Li: Okay, then, what's next?

510

00:55:48.070 --> 00:55:52.760

Didong Li: Okay? The next topic is, it's fairly simple.

511

00:55:55.790 --> 00:55:57.420

Didong Li: It's grammar correction.

512

00:55:57.900 --> 00:56:01.869

Didong Li: although this is a simple paragraph. Let's read it together here.

513

00:56:04.600 --> 00:56:06.100

Didong Li: If you find a

514

00:56:06.440 --> 00:56:08.350

Didong Li: mistake or table.

515

00:56:08.610 --> 00:56:09.550

Didong Li: Let me know

516

00:56:49.410 --> 00:56:50.220

Didong Li: in the past

517

00:56:52.330 --> 00:56:54.199

Didong Li: where which line please.

518

00:56:56.784 --> 00:57:00.979

Didong Li: Researchers fails, no matter who fails to fail.

519

00:57:03.570 --> 00:57:08.090

Didong Li: which has been.

520

00:57:08.610 --> 00:57:10.920

Didong Li: it's high. Right? Okay, yeah.

521

00:57:11.040 --> 00:57:11.720

Didong Li: Cool.

522

00:57:12.110 --> 00:57:13.060

Didong Li: Last.

523

00:57:14.591 --> 00:57:19.599

Didong Li: this method for life. Okay.

524

00:57:19.700 --> 00:57:21.060

Didong Li: rolls also evolved.

525

00:57:23.790 --> 00:57:25.290

Didong Li: It's nice.

526

00:57:25.450 --> 00:57:27.899

Didong Li: including jeans. Open

527

00:57:28.150 --> 00:57:30.060

Didong Li: tuck normally.

528

00:57:30.070 --> 00:57:31.470

Didong Li: where? Where? Which line?

529

00:57:34.470 --> 00:57:37.700

Didong Li: Oh, okay, okay. This happens.

530

00:57:41.400 --> 00:57:47.819

Didong Li: The one about it, the technique about it. One of the most commonly used techniques, okay.

531

00:57:54.830 --> 00:57:58.660

Didong Li: actually does not have to stay in the

532

00:57:58.700 --> 00:58:01.920

Didong Li: so originally, which allows to understand

533

00:58:01.980 --> 00:58:03.970

Didong Li: it's not the mythical works

534

00:58:04.290 --> 00:58:06.740

Didong Li: which allows food.

535

00:58:07.320 --> 00:58:09.920

Didong Li: Allow someone to understand. Right?

536

00:58:10.250 --> 00:58:11.010

Didong Li: Okay.

537

00:58:14.960 --> 00:58:15.970

Didong Li: nothing else.

538

00:58:16.110 --> 00:58:17.550

Didong Li: Physical, too long

539

00:58:17.890 --> 00:58:20.930

Didong Li: multiplied, should be

540

00:58:21.120 --> 00:58:21.870

Didong Li: leading to

541

00:58:22.110 --> 00:58:23.410

Didong Li: equity results

542

00:58:27.030 --> 00:58:27.740

Didong Li: cool.

543

00:58:33.540 --> 00:58:36.530

Didong Li: It looked like a real parallel.

544

00:58:36.730 --> 00:58:38.239

Didong Li: Fast food owners.

545

00:58:42.380 --> 00:58:43.210

Didong Li: Me.

546

00:58:44.160 --> 00:58:45.170

Didong Li: that's good.

547

00:58:45.940 --> 00:58:48.049

Didong Li: What's a swim by touching here

548

00:58:48.720 --> 00:58:52.949

Didong Li: last? Can you read a paragraph about statistics and based on titles.

549

00:58:53.727 --> 00:59:03.929

Didong Li: This looks like pretty rare. The mistakes are very common, at least for me, English is done, not my 1st language sometimes made us the same mistake.

550

00:59:05.300 --> 00:59:08.829

Didong Li: okay? So you can find this paragraph in the Google, Doc.

551

00:59:09.536 --> 00:59:11.800

Didong Li: let's try this together.

552

00:59:13.010 --> 00:59:16.099

Didong Li: Check whether Gbt can catch this.

553

00:59:17.500 --> 00:59:21.090

Didong Li: Okay, so grammar correction.

554

00:59:25.770 --> 00:59:31.350

Didong Li: So so this is a little bit tricky. If you just ask this to do grammar correction, it may

555

00:59:31.550 --> 00:59:35.440

Didong Li: give you a new paragraph with the grammar fixed.

556

00:59:35.610 --> 00:59:42.800

Didong Li: but you want to know where the the the table is? Okay. So we can do something like, can you

557

00:59:43.140 --> 00:59:44.390

Didong Li: correct

558

00:59:45.020 --> 00:59:46.110

Didong Li: Graham, or

559

00:59:50.480 --> 00:59:57.560

Didong Li: and both face the changes or tables, so I'll ask it to both face

560

00:59:58.590 --> 01:00:00.569

Didong Li: those so that you can check.

561

01:00:02.720 --> 01:00:03.400

Didong Li: I mean.

562

01:00:03.580 --> 01:00:06.309

Didong Li: check our answers. Compare with this.

563

01:00:07.020 --> 01:00:07.970

Didong Li: Okay.

564

01:00:08.490 --> 01:00:11.170

Didong Li: the 1st one I have turned to Hans

565

01:00:11.930 --> 01:00:13.000

Didong Li: just here.

566

01:00:13.410 --> 01:00:15.690

Didong Li: This method provide

567

01:00:16.400 --> 01:00:17.550

Didong Li: techniques.

568

01:00:17.820 --> 01:00:21.540

Didong Li: Allow researchers to understand fail

569

01:00:21.710 --> 01:00:22.980

Didong Li: these techniques.

570

01:00:23.480 --> 01:00:24.560

Didong Li: It seems

571

01:00:25.670 --> 01:00:26.819

Didong Li: working well.

572

01:00:27.830 --> 01:00:29.709

Didong Li: Anything missing here.

573

01:00:34.960 --> 01:00:37.159

Didong Li: Oh, sorry I

574

01:00:37.300 --> 01:00:38.790

Didong Li: how to zoom in.

575

01:01:05.050 --> 01:01:05.790

Didong Li: Good.

576

01:01:07.350 --> 01:01:09.879

Didong Li: Yeah, I literally do this all the time.

577

01:01:11.230 --> 01:01:13.280

Didong Li: And actually, I do have a

578

01:01:15.370 --> 01:01:19.590

Didong Li: we. We already write paper together on our deal, not sure whether we use it.

579

01:01:20.320 --> 01:01:21.540

Didong Li: It's a

580

01:01:22.890 --> 01:01:24.850

Didong Li: and yeah, I'm not selling this.

581

01:01:26.710 --> 01:01:28.999

Didong Li: Oh, I don't know what he's trying to show you.

582

01:01:30.060 --> 01:01:31.820

Didong Li: Maybe I'll go back to this.

583

01:01:34.110 --> 01:01:39.749

Didong Li: So there's a building software called Red 4. It can detect your grammar

584

01:01:39.930 --> 01:01:42.029

Didong Li: like tables. Whatever. Give you suggestions

585

01:01:42.280 --> 01:01:44.590

Didong Li: automatically. But it's not free.

586

01:01:44.710 --> 01:01:47.520

Didong Li: I think some universities have access to this.

587

01:01:47.660 --> 01:01:59.529

Didong Li: But anyway, you can. You can always use Gbt, which is free. So after you write a paper, you go through this to catch typos that's allowed by most journals. We will talk about the policy later.

588

01:01:59.610 --> 01:02:03.859

Didong Li: This is a very helpful use of tragically, in my point of view.

589

01:02:06.660 --> 01:02:10.280

Didong Li: any questions to others, please.

590

01:02:10.530 --> 01:02:12.250

Didong Li: So this display.

591

01:02:12.910 --> 01:02:14.040

Didong Li: or live

592

01:02:15.520 --> 01:02:17.060

Didong Li: or leave.

593

01:02:17.320 --> 01:02:22.120

Didong Li: or if it's it's free, you can. You can write papers with your collaborators

594

01:02:22.430 --> 01:02:23.360

Didong Li: together.

595

01:02:26.250 --> 01:02:28.860

Didong Li: Any other

596

01:02:30.010 --> 01:02:32.290

Didong Li: right for

597

01:02:33.330 --> 01:02:34.480

Didong Li: rightful.

598

01:02:40.020 --> 01:02:46.669

Didong Li: Yeah, it's it's not. It's not super smart, but it's it's sometimes helpful.

599

01:02:47.150 --> 01:02:49.191

Didong Li: It can make it can.

600

01:02:51.000 --> 01:02:54.590

Didong Li: it's a little bit embarrassing. But let me show you a paper

601

01:02:55.590 --> 01:02:58.037

Didong Li: I don't know which one which one will show.

602

01:03:03.560 --> 01:03:06.529

Didong Li: It's on tech latex compeller. Basically.

603

01:03:12.790 --> 01:03:14.390

Didong Li: some features are

604

01:03:15.670 --> 01:03:20.509

Didong Li: like a lot of universities have to tell you.

605

01:03:22.010 --> 01:03:24.989

Didong Li: Yeah, you can see this is a, this is real paper.

606

01:03:25.100 --> 01:03:26.480

Didong Li: And the

607

01:03:26.590 --> 01:03:28.649

Didong Li: this is so embarrassing.

608

01:03:33.030 --> 01:03:36.490

Didong Li: Well, this this our white line means there's a

609

01:03:37.140 --> 01:03:38.700

Didong Li: potential issue.

610

01:03:41.050 --> 01:03:44.499

Didong Li: Yes, yes, you have to check. You know, this has.

611

01:03:44.720 --> 01:03:50.280

Didong Li: you know, red voice, 94%. Sure they should be

612

01:03:50.310 --> 01:03:51.410

Didong Li: another word.

613

01:03:51.750 --> 01:03:54.870

Didong Li: It's it's not. It's not always working. Yeah, I agree with it.

614

01:03:55.050 --> 01:03:55.950

Didong Li: Of.

615

01:03:57.010 --> 01:03:58.469

Didong Li: okay. So

616

01:03:59.070 --> 01:04:01.489

Didong Li: let's go back. There's any question about this.

617

01:04:03.490 --> 01:04:09.339

Didong Li: Okay. But let's do a 5Â min break, and next we will move on to Latex and coding.

618

01:04:09.950 --> 01:04:12.849

Didong Li: which is pretty challenging. Motivity is not really good at

619

01:04:14.630 --> 01:04:16.470

Didong Li: okay, let me.

620

01:04:17.700 --> 01:04:20.080

Didong Li: Should I stop sharing. I'm recording or

621

01:04:20.520 --> 01:04:21.240

Didong Li: keep

622

01:04:23.070 --> 01:04:25.710

Didong Li: okay. Cool. Yeah. Cool, cool, cool.

623

01:04:26.570 --> 01:04:27.430

Didong Li: possible.

624

01:04:30.100 --> 01:04:30.860

Didong Li: Stop it.

625

01:04:33.930 --> 01:04:36.510

Didong Li: I can pause it. Yes.

626

01:04:36.660 --> 01:04:37.520

Didong Li: Okay.

627

01:04:38.190 --> 01:04:47.489

Didong Li: okay, cool. Okay. Welcome back the remaining time. I'll talk about some later kind of coding tasks.

628

01:04:48.780 --> 01:04:51.959

Didong Li: The 1st one needs to generate tables

629

01:04:52.850 --> 01:04:54.219

Didong Li: for a later.

630

01:04:56.820 --> 01:04:59.039

Didong Li: It was very convenient to that

631

01:04:59.320 --> 01:05:06.020

Didong Li: tax in Latex. You don't need to code at all, but sometimes we deal with some complex tables and figures.

632

01:05:06.070 --> 01:05:10.110

Didong Li: I personally struggle a lot in making these figures and tables.

633

01:05:10.450 --> 01:05:12.709

Didong Li: especially if the table is pretty

634

01:05:12.850 --> 01:05:24.639

Didong Li: non standard. So we'll discuss 2 user cases here. One is to generate tables from some very detailed prompts. For example.

635

01:05:24.760 --> 01:05:27.910

Didong Li: here we ask Gpt to generate a table.

636

01:05:28.140 --> 01:05:31.867

Didong Li: payrolls for columns with different colors, like

637

01:05:32.630 --> 01:05:34.920

Didong Li: spending columns for words

638

01:05:35.000 --> 01:05:41.750

Didong Li: cells something like that. You can feel free to read it. And this tax is also in the Google Doc.

639

01:05:42.710 --> 01:05:46.420

Didong Li: And let's ask Gpt to generate a little code for this table.

640

01:05:48.100 --> 01:05:50.660

Didong Li: So let's try.

641

01:06:01.610 --> 01:06:02.790

Didong Li: Well, I'm good.

642

01:06:03.420 --> 01:06:06.480

Didong Li: Create a new chat in case

643

01:06:07.530 --> 01:06:12.940

Didong Li: literally copied here. General little code for table with the following specifications.

644

01:06:19.650 --> 01:06:23.159

Didong Li: we always, I'm not sure it works or not. That's fine.

645

01:06:42.510 --> 01:06:43.420

Didong Li: Okay?

646

01:06:43.590 --> 01:06:48.239

Didong Li: Looks like later code. And let's compile it here.

647

01:06:55.680 --> 01:06:56.360

Didong Li: Oops.

648

01:06:59.430 --> 01:07:00.500

Didong Li: Okay.

649

01:07:02.030 --> 01:07:03.609

Didong Li: what are the arrows here?

650

01:07:04.910 --> 01:07:07.750

Didong Li: Okay. Undefined color light blue.

651

01:07:13.140 --> 01:07:14.380

Didong Li: Where is the

652

01:07:16.825 --> 01:07:17.609

Didong Li: okay?

653

01:07:18.120 --> 01:07:22.149

Didong Li: How about we change it to blue for now, so that it can compile at least.

654

01:07:24.270 --> 01:07:25.210

Didong Li: Okay.

655

01:07:27.660 --> 01:07:29.310

Didong Li: so what's the

656

01:07:29.450 --> 01:07:31.699

Didong Li: what are the requirements. I can't remember

657

01:07:32.330 --> 01:07:34.360

Didong Li: 10 or 5 columns.

658

01:07:34.720 --> 01:07:36.410

Didong Li: Sense. Okay.

659

01:07:36.590 --> 01:07:41.310

Didong Li: 1st column in gray is like the 3rd column in blue, although the color is not exact.

660

01:07:41.630 --> 01:07:43.350

Didong Li: You can play with it later.

661

01:07:43.580 --> 01:07:44.944

Didong Li: There are some

662

01:07:45.560 --> 01:07:49.070

Didong Li: merge styles, multirole spanning 3 rows.

663

01:07:49.470 --> 01:07:51.689

Didong Li: no, 2 rows fanning all columns.

664

01:07:52.272 --> 01:07:53.840

Didong Li: Seems seems working.

665

01:07:57.120 --> 01:07:58.700

Didong Li: was simple, and questions.

666

01:08:00.100 --> 01:08:01.140

Didong Li: Work for you.

667

01:08:01.380 --> 01:08:02.610

Didong Li: Does it work for you

668

01:08:04.460 --> 01:08:06.400

Didong Li: is a very different

669

01:08:07.640 --> 01:08:08.770

Didong Li: okay?

670

01:08:08.940 --> 01:08:16.790

Didong Li: Maybe we didn't specify the location of the Morse or standing south.

671

01:08:17.689 --> 01:08:23.674

Didong Li: It's okay. I'll show you one example where I'm pretty confident that the result will be very different. Okay.

672

01:08:24.590 --> 01:08:28.450

Didong Li: so this is the user case, especially useful when you

673

01:08:28.490 --> 01:08:30.831

Didong Li: you know. What does the

674

01:08:32.930 --> 01:08:35.340

Didong Li: what does the figure exact?

675

01:08:35.439 --> 01:08:37.590

Didong Li: What does the table exactly look like?

676

01:08:38.939 --> 01:08:40.340

Didong Li: But sometimes

677

01:08:41.330 --> 01:08:42.200

Didong Li: you!

678

01:08:42.220 --> 01:08:46.309

Didong Li: This is why I generate from when I prepare for the slides. It's pretty

679

01:08:46.350 --> 01:08:53.549

Didong Li: similar, but the difference is, there's another multi-row cell here. You can specify the location and number of motels.

680

01:08:53.680 --> 01:08:54.910

Didong Li: Sometimes.

681

01:08:55.229 --> 01:08:56.879

Didong Li: you know you find a.

682

01:08:57.060 --> 01:09:18.440

Didong Li: The table that you you feel is is great, is a beautiful table from other people's other people's paper. You want to ask Gpd. To generate the same one for you, for example, this is a this is a table from from my paper. It's not beautiful, but it's just example. You know. I want to show you some example. The hardest part is there are some bold face numbers.

683

01:09:18.640 --> 01:09:29.580

Didong Li: Usually, when you present this result, you both face the best performance. So we can ask Cbd to do to do this. If you want, you can try. This is a screenshot.

684

01:09:30.270 --> 01:09:31.759

Didong Li: you know, Google, Doc.

685

01:09:31.830 --> 01:09:36.210

Didong Li: And you can ask, can you generate a code for figure for table like this?

686

01:09:38.359 --> 01:09:42.820

Didong Li: Okay, so this is my prompt. I won't repeat it here. Generate a code for this table

687

01:09:44.689 --> 01:09:50.890

Didong Li: so that it can generate exact table with exactly the same number, same row columns, and both face the same cells.

688

01:09:58.570 --> 01:09:59.170

Didong Li: Course.

689

01:10:03.210 --> 01:10:04.100

Didong Li: yeah.

690

01:10:13.660 --> 01:10:15.630

Didong Li: Now, I use latex code.

691

01:10:15.730 --> 01:10:16.750

Didong Li: So

692

01:10:16.860 --> 01:10:21.940

Didong Li: fact that when I was in school

693

01:10:22.010 --> 01:10:27.670

Didong Li: oh, it's not between

694

01:10:27.710 --> 01:10:30.380

Didong Li: paperwork and integrations.

695

01:10:31.990 --> 01:10:34.320

Didong Li: rather, whether you want

696

01:10:35.980 --> 01:10:37.410

Didong Li: not that easy to

697

01:10:38.680 --> 01:10:39.939

Didong Li: has a lot of equations

698

01:10:45.760 --> 01:10:46.980

Didong Li: himself.

699

01:10:48.019 --> 01:10:52.319

Didong Li: So in writing you all.

700

01:10:53.180 --> 01:10:54.520

Didong Li: it's a decent tax paper.

701

01:10:54.850 --> 01:10:55.780

Didong Li: Excuse me.

702

01:10:59.140 --> 01:11:03.220

Didong Li: and it doesn't prove anything right.

703

01:11:13.310 --> 01:11:16.100

Didong Li: I'll share those slides later. So no worries.

704

01:11:21.640 --> 01:11:23.150

Didong Li: So how

705

01:11:23.590 --> 01:11:25.490

Didong Li: positive performance

706

01:11:27.450 --> 01:11:29.179

Didong Li: didn't work didn't try.

707

01:11:29.390 --> 01:11:33.370

Didong Li: Okay, I'll show you the result I got last time.

708

01:11:34.490 --> 01:11:39.449

Didong Li: This is one. This is the compiled version of the table, from from

709

01:11:39.780 --> 01:11:45.199

Didong Li: from Gpt. But the overall structure seems okay. But there are some problems.

710

01:11:46.160 --> 01:11:47.040

Didong Li: Well.

711

01:11:47.180 --> 01:11:49.070

Didong Li: this is open.

712

01:11:49.770 --> 01:11:53.679

Didong Li: because there's no numbers here to be almost confused.

713

01:11:53.710 --> 01:12:00.519

Didong Li: And also, if you look at the bold face numbers, it's not really correct. For example, this one should not be bold. Face

714

01:12:00.600 --> 01:12:01.740

Didong Li: this one.

715

01:12:02.440 --> 01:12:05.819

Didong Li: this one. There are some mistake in the boat phase.

716

01:12:06.800 --> 01:12:07.600

Didong Li: Hmm!

717

01:12:08.080 --> 01:12:09.210

Didong Li: How to fix it.

718

01:12:10.000 --> 01:12:13.080

Didong Li: and fortunately there are some

719

01:12:13.100 --> 01:12:15.180

Didong Li: fine-tune tragic achieve

720

01:12:15.450 --> 01:12:16.780

Didong Li: for little red Hats

721

01:12:17.040 --> 01:12:18.270

Didong Li: generation.

722

01:12:19.980 --> 01:12:36.380

Didong Li: for example, there is one called Latex Transformer. I will show you some other examples also related to the question from the audience. There are some like apps available on the still in Chat Gp website. They are fine tuned for different purposes. They are really

723

01:12:36.800 --> 01:12:50.169

Didong Li: stronger in some very detailed task than Chat Gpt. It's still part of Chat Gpt is more like a a app there. So this app, I showed is is called Latex Transformer. It's just

724

01:12:50.430 --> 01:12:59.140

Didong Li: designed to generate latex code. So ask the same question to this app. I'll show you how to do it in a sec. And the result is

725

01:12:59.380 --> 01:13:02.639

Didong Li: much better. First, st this is closed

726

01:13:02.750 --> 01:13:06.470

Didong Li: and the both phase are correct this time, as as you can check

727

01:13:06.680 --> 01:13:08.180

Didong Li: and how to use it.

728

01:13:09.700 --> 01:13:12.660

Didong Li: Let's go to try Gpt, website.

729

01:13:15.220 --> 01:13:16.750

Didong Li: Okay, you click.

730

01:13:16.920 --> 01:13:19.720

Didong Li: We explore gpts on the left.

731

01:13:20.830 --> 01:13:22.670

Didong Li: you'll see that option.

732

01:13:23.170 --> 01:13:26.260

Didong Li: Okay? Then you search for

733

01:13:26.570 --> 01:13:27.790

Didong Li: later.

734

01:13:30.960 --> 01:13:36.980

Didong Li: And the 1st one is latex transformer. You can see a lot of apps related to late Latex. Cowper, Latex.

735

01:13:37.170 --> 01:13:40.860

Didong Li: Cheap, late export mass for Malaysia, and this is more like

736

01:13:40.990 --> 01:13:46.699

Didong Li: more good at like massive questions. Text related color, picture blah blah blah.

737

01:13:47.070 --> 01:13:52.660

Didong Li: and you can still number of uses after after each app question

738

01:13:56.380 --> 01:13:57.550

Didong Li: the

739

01:13:59.430 --> 01:14:02.190

Didong Li: Oh, you just click explore gpts

740

01:14:04.270 --> 01:14:05.370

Didong Li: on the left.

741

01:14:05.700 --> 01:14:07.220

Didong Li: Funded. Okay.

742

01:14:07.250 --> 01:14:10.799

Didong Li: I will search for whatever like and find this. You click on this

743

01:14:11.240 --> 01:14:12.669

Didong Li: and we'll create

744

01:14:12.980 --> 01:14:15.789

Didong Li: start chat. You know. There, there are regions.

745

01:14:15.930 --> 01:14:17.350

Didong Li: capability.

746

01:14:17.500 --> 01:14:19.899

Didong Li: the congenitary images from another.

747

01:14:20.435 --> 01:14:26.019

Didong Li: It's also it's called Deli, from Openai. It's more good at images

748

01:14:26.120 --> 01:14:30.129

Didong Li: and code interpreter data analysis. And you click, start chat.

749

01:14:31.150 --> 01:14:34.629

Didong Li: And then it's nothing different. It's just another chat window.

750

01:14:35.350 --> 01:14:40.660

Didong Li: But you are really chatting with this app, not the general chat. Gpt. And you can ask, like

751

01:14:41.060 --> 01:14:42.850

Didong Li: Shannerine.

752

01:14:45.060 --> 01:14:51.920

Didong Li: wait a code for the table. I won't do it here. It's due to the time limit.

753

01:14:52.350 --> 01:14:55.449

Didong Li: So the result I got from this is

754

01:14:56.530 --> 01:14:59.030

Didong Li: apparently better than the naive version.

755

01:14:59.710 --> 01:15:03.199

Didong Li: because it's fine-tuned for Latex code. Generation.

756

01:15:06.615 --> 01:15:07.270

Didong Li: Questions.

757

01:15:10.025 --> 01:15:10.440

Didong Li: This.

758

01:15:10.900 --> 01:15:14.659

Didong Li: I think mathematics is the best of.

759

01:15:15.220 --> 01:15:17.599

Didong Li: not a generator, not a generator.

760

01:15:19.030 --> 01:15:21.810

Didong Li: If you will check the Math Peaks

761

01:15:23.540 --> 01:15:27.810

Didong Li: is very better than the chat that you can have other software. Yeah.

762

01:15:27.930 --> 01:15:31.370

Didong Li: yeah. But there is no in the list of charges between my fixed

763

01:15:32.530 --> 01:15:36.169

Didong Li: yeah, May. Maybe there is one, or you can create one.

764

01:15:36.260 --> 01:15:38.270

Didong Li: you know. Create a app.

765

01:15:40.310 --> 01:15:48.950

Didong Li: Yes, there is an app. Okay, it's not. I thought, it's it's very, very good. Okay. I can generate, for example.

766

01:15:49.190 --> 01:15:58.360

Didong Li: 1,000 page by one page in like immediately to like finish, I think, is the best, but it was not in the list of tragedy.

767

01:15:58.660 --> 01:15:59.939

Didong Li: Oh, really.

768

01:16:00.170 --> 01:16:01.120

Didong Li: there's no

769

01:16:01.724 --> 01:16:11.719

Didong Li: oh, I see, I see. Okay. But to be clear, you know, you have to pay some size if you use it

770

01:16:12.030 --> 01:16:14.460

Didong Li: our day.

771

01:16:14.480 --> 01:16:18.699

Didong Li: Yeah, yeah, I see. I see. Okay, good to know if you want, you can share with us.

772

01:16:19.420 --> 01:16:20.350

Didong Li: Thanks.

773

01:16:22.030 --> 01:16:24.260

Didong Li: Okay, so this is why app.

774

01:16:25.150 --> 01:16:34.859

Didong Li: and then let's move on to plots. And this time we have the detailed prompts like we want a multi panel plot 2 by 2.

775

01:16:35.759 --> 01:16:38.409

Didong Li: You know, labeled as Abcd.

776

01:16:40.653 --> 01:16:45.390

Didong Li: Each panel has a subtitle. There is a main title.

777

01:16:47.040 --> 01:16:49.799

Didong Li: you know, evenly spaced, spaced, and aligned.

778

01:16:50.440 --> 01:16:56.279

Didong Li: and the title is, is like this. This is also a very detailed prompt to generate about.

779

01:16:58.560 --> 01:17:03.560

Didong Li: include a label for referencing the multi-panel about document.

780

01:17:03.690 --> 01:17:06.520

Didong Li: This is pretty standard when we read papers

781

01:17:06.750 --> 01:17:08.999

Didong Li: and explain the later code

782

01:17:09.840 --> 01:17:11.600

Didong Li: or make it readable.

783

01:17:12.290 --> 01:17:19.460

Didong Li: So this is the detailed prompt. You can find this in the Google Doc, and again, due to time limit. I'll skip it here. It doesn't work.

784

01:17:20.230 --> 01:17:21.240

Didong Li: basically.

785

01:17:22.770 --> 01:17:25.719

Didong Li: So maybe we have to go back to

786

01:17:25.980 --> 01:17:27.639

Didong Li: the late hack transformer

787

01:17:28.140 --> 01:17:33.219

Didong Li: is is more tailored for later code. It works really well.

788

01:17:33.850 --> 01:17:36.279

Didong Li: at least to generate a 2 by 2. Figure.

789

01:17:36.670 --> 01:17:40.188

Didong Li: I'll skip one more example due to time limit.

790

01:17:40.880 --> 01:17:42.470

Didong Li: I go to coding part.

791

01:17:45.733 --> 01:17:50.279

Didong Li: Okay? So we want to ask Gpu to explain our code.

792

01:17:50.520 --> 01:17:53.840

Didong Li: This is a very simple chunk of our code for

793

01:17:53.910 --> 01:17:57.890

Didong Li: generalized video model. Our function is called Jrm.

794

01:17:59.780 --> 01:18:03.770

Didong Li: This is helpful when you know you really understand the

795

01:18:04.260 --> 01:18:12.860

Didong Li: our our language. But you are familiar with another language. For example, I want to make sure. What does each each row or each sentence mean?

796

01:18:13.050 --> 01:18:15.129

Didong Li: And you just literally copy this

797

01:18:15.210 --> 01:18:19.350

Didong Li: prompt to to tactability, explain the following, our code line by line.

798

01:18:19.880 --> 01:18:21.170

Didong Li: and

799

01:18:22.210 --> 01:18:25.039

Didong Li: maybe you can do it now, or you can do it later.

800

01:18:25.270 --> 01:18:30.019

Didong Li: As of time, constraint. So my experience is is doing really really well explaining

801

01:18:30.070 --> 01:18:32.640

Didong Li: existing our code, if it makes sense

802

01:18:35.120 --> 01:18:38.250

Didong Li: or we can generate our code for simple tasks.

803

01:18:38.360 --> 01:18:40.220

Didong Li: for example.

804

01:18:40.240 --> 01:18:50.309

Didong Li: But this is a very detailed, prompt I I designed from chatting fee. Actually, to use the building iris data set, perform a summary analysis.

805

01:18:50.660 --> 01:18:52.570

Didong Li: conduct Pca

806

01:18:52.800 --> 01:19:01.000

Didong Li: visualize the Pca. Result. These scatter plots colored by species. There are 3 species in the data set. Create a pair plot.

807

01:19:01.710 --> 01:19:04.580

Didong Li: you know. Explain your code clearly.

808

01:19:05.440 --> 01:19:08.249

Didong Li: I think I have the result in a Google, Doc.

809

01:19:08.710 --> 01:19:09.750

Didong Li: $4,

810

01:19:14.440 --> 01:19:15.620

Didong Li: or maybe not.

811

01:19:16.200 --> 01:19:17.360

Didong Li: So

812

01:19:17.520 --> 01:19:19.629

Didong Li: I'll get in this works

813

01:19:19.770 --> 01:19:22.160

Didong Li: really? Well, if you want, you can try this.

814

01:19:22.200 --> 01:19:30.050

Didong Li: But if you want to ask you to write a very complicated package. I don't think it works but for this type of simple task this works really well.

815

01:19:30.950 --> 01:19:32.090

Didong Li: The bog

816

01:19:32.500 --> 01:19:38.619

Didong Li: this will be challenging, so I won't ask you to debug here. There are some very

817

01:19:40.020 --> 01:19:54.189

Didong Li: there are some bugs that hard to detect. If you, if you read it here again, I generate this code, this box by chat Vt, ask me to generate some code, this box that are like hard to identify. There are some like a

818

01:19:54.450 --> 01:20:12.329

Didong Li: the scale. There's a dot here. Actually, the.is amazing. It's just some like grammar in R or some like a capitalized species hard to catch. We make the same mistake when we code that ourselves. And so at least when when I try this, Gbd works really well to detect this box

819

01:20:13.400 --> 01:20:23.629

Didong Li: message translation is also pretty simple, especially when you deal with very commonly used languages like Python. But, according to my experience, it's not really good at Sas.

820

01:20:24.000 --> 01:20:29.590

Didong Li: probably because of the training data. So Tb coding side was trained on Github

821

01:20:30.273 --> 01:20:41.399

Didong Li: Because Microsoft is one of the donors of of Openai. So they have access to Github, and if you look at Github Repo, how many packages are in sauce?

822

01:20:41.710 --> 01:20:52.369

Didong Li: The proportion is pretty small. That's why it's not really powerful there, but if you do it for R or python, it's it's really worked really well here. I literally ask Gbg. To translate

823

01:20:52.670 --> 01:20:58.809

Didong Li: Iris example to python. It works really well, I ran it, and it it works

824

01:21:00.581 --> 01:21:06.059

Didong Li: so for the remaining time I would like to discuss something different. For example.

825

01:21:06.080 --> 01:21:09.710

Didong Li: as I mentioned at the very beginning. Gpt is

826

01:21:09.930 --> 01:21:19.690

Didong Li: this model is not trained for any specific task and Chat Gpt is a conversational expert trained, based on Gpt.

827

01:21:19.770 --> 01:21:25.980

Didong Li: And this open air Codecs is another expert train for coding.

828

01:21:26.170 --> 01:21:28.820

Didong Li: If you're if you go with just

829

01:21:28.880 --> 01:21:31.909

Didong Li: red calls, this is better than tragedy.

830

01:21:32.080 --> 01:21:46.129

Didong Li: because this is trained just for coding. It's called Codex. It was born long before Chat Gpt cause. In some sense coding is easier than conversation, because it's more structured. It's it's more follows some certain logics.

831

01:21:47.100 --> 01:21:48.200

Didong Li: Oh.

832

01:21:48.290 --> 01:21:49.380

Didong Li: or.

833

01:21:49.440 --> 01:21:52.719

Didong Li: as we have seen sometimes you have other

834

01:21:53.450 --> 01:22:05.439

Didong Li: other goals if you want to do latex generation. We have seen the latex transformer, but sometimes you want to do something else. If you click the explore gpts, you can see different categories.

835

01:22:05.620 --> 01:22:26.410

Didong Li: These are some featured topics for different purposes. Something related to our research or writing. If you click on writing, you see this right for me, humanized AI. One complaint is, if you use Cbt to polish your paper, and you use some really weird words that are never used a human being.

836

01:22:26.480 --> 01:22:32.349

Didong Li: And you can use this humanized AI to make it more like a human written text.

837

01:22:33.050 --> 01:22:35.539

Didong Li: Again, there are some other apps you want to try.

838

01:22:35.560 --> 01:22:36.620

Didong Li: Oh.

839

01:22:37.280 --> 01:22:47.330

Didong Li: research and analysis. There's a scholar. Gpt, scholar AI. Can. It can do search on Google scholar, or some other database.

840

01:22:47.380 --> 01:22:48.730

Didong Li: These are

841

01:22:48.780 --> 01:22:58.050

Didong Li: this one, you know. The is more like a mathematica or style. It can do some math computation or abstract symbolic computation

842

01:22:58.870 --> 01:23:12.239

Didong Li: or education use so relate to a question from the audience. There are some math tutors on math solver. This is pretty strong in solving, like high school level math problems, maybe even undergrad students.

843

01:23:12.370 --> 01:23:27.019

Didong Li: but not really good at solving, like Phd level math problems. These are for really education use. I have some colleagues who use this almost every day when their kids ask questions say, high school or middle school math questions.

844

01:23:27.690 --> 01:23:37.719

Didong Li: and it's pretty accurate, actually, or programming. This is a copilot where you write a code. This the app together, like python. If you search for R, you'll see some apps.

845

01:23:37.950 --> 01:23:45.880

Didong Li: and if you are not satisfied by these apps, you can build your own gpties, so I'll show you quickly how to do it here.

846

01:23:46.783 --> 01:23:48.270

Didong Li: You click.

847

01:23:49.566 --> 01:23:51.120

Didong Li: explore Tps

848

01:23:52.170 --> 01:23:54.859

Didong Li: and click my GPS,

849

01:23:56.270 --> 01:23:59.289

Didong Li: and you pick, create a Gpt.

850

01:23:59.710 --> 01:24:19.680

Didong Li: and this is a builder you wanna communicate with. Imagine this is a person human being you want to communicate with this person. I want to use this create app for Latex code generation. I want to create app for a a Pdf. Or create a a app for math problem or mess solver.

851

01:24:19.730 --> 01:24:30.759

Didong Li: And you can upload a lot of Pdfs. Here, for example, you can upload all your previous papers. You can ask, can you write a draft meaning my style, my writing style?

852

01:24:30.830 --> 01:24:36.339

Didong Li: You know, it allows for more training data here is more targeted for some specific task.

853

01:24:36.400 --> 01:24:41.670

Didong Li: And if your your app is working really well, you wanna you can make it public.

854

01:24:41.890 --> 01:24:48.369

Didong Li: and I think check the opening app will pay you if it's like it's like a

855

01:24:48.420 --> 01:25:04.269

Didong Li: app store or android store. If it's very popular, you you get some money from this. From this app. It really depends. You can see the Latex, late transformer was downloaded like 25,000 times every day. They will be paid because of this app.

856

01:25:04.390 --> 01:25:06.010

Didong Li: and then so

857

01:25:08.050 --> 01:25:11.860

Didong Li: so these are 2 apps I built recently. One is

858

01:25:11.950 --> 01:25:29.030

Didong Li: to prepare for my teaching is called Advanced machine Learning Bells. At 7, 74. You can see the description how to prepare teacher materials for PC course in advance machine learning, and the 1st app is used to prepare slides for today. I call this workshop preparation assistant.

859

01:25:29.150 --> 01:25:36.290

Didong Li: you know, for research working, writing workshop. That's how I created those mock tags for grammar correction

860

01:25:36.330 --> 01:25:39.999

Didong Li: with some real tables, or I create those

861

01:25:40.250 --> 01:25:42.890

Didong Li: our code with real tables.

862

01:25:42.940 --> 01:25:48.650

Didong Li: So if you want you can, you can create your own app for some specific tasks.

863

01:25:49.730 --> 01:25:53.390

Didong Li: So at the end, I would like to discuss some limitations.

864

01:25:53.420 --> 01:25:54.740

Didong Li: especially

865

01:25:55.207 --> 01:26:14.189

Didong Li: this is a growing area. There are a lot of policies, a lot of uncertainty when you use Gpt. For example, I copied from the policy from Europe's a Machine Learning Conference. They have very detailed policy about the usage of license models. So basically, you can use it

866

01:26:14.290 --> 01:26:20.740

Didong Li: to polish your paper, but it's dangerous if you read it from the thread, you actually read it entire paper.

867

01:26:20.750 --> 01:26:38.100

Didong Li: That's sometimes not a lot. And no matter what you generate from the large model, you should be responsible to what you wrote down in your paper, not the large model. There are some like a conversation about this, even from the government side. This is a open area.

868

01:26:38.480 --> 01:26:41.030

Didong Li: I just copy the policy from new ribs.

869

01:26:41.050 --> 01:26:59.260

Didong Li: And so at Unc. Papu hill. When we teach our course. You know students, they sometimes use Gpt to solve their homework problems, to write their report, their final project, report something like that. And so this is a policy in my course and in different courses. There are different policies.

870

01:26:59.260 --> 01:27:15.140

Didong Li: So I would say, if you use Gbt to write something, either a research paper or homework problem or even technical report. You better check the policy. You know it's it's very dangerous if you use it without aware of the the policy against it. You know the

871

01:27:15.220 --> 01:27:19.799

Didong Li: you have. You're you have to be responsible to whatever you see the

872

01:27:20.020 --> 01:27:21.459

Didong Li: in your documents.

873

01:27:21.750 --> 01:27:31.419

Didong Li: So I guess it's time to wrap up today. We briefly talk about the history of categt, and we use it to summarize a research paper

874

01:27:31.430 --> 01:27:37.279

Didong Li: to write an outline, do grammar correction, generate code for tables and figures?

875

01:27:37.832 --> 01:27:40.509

Didong Li: Draws a little bit in the coding

876

01:27:40.560 --> 01:27:45.790

Didong Li: chapter, explain our code, generate our codes, debug and translate.

877

01:27:45.930 --> 01:28:00.020

Didong Li: and we briefly talk about extensions to use other apps, or to build your own apps, or, if you are really interested, you can use other large models, say llama, which is open source to fine tune by yourself for your domain tasks.

878

01:28:01.060 --> 01:28:04.749

Didong Li: for example, you have some electronic house record data with

879

01:28:05.030 --> 01:28:10.769

Didong Li: often subject to some privacy policy. You can't really upload the data to catch Gbt.

880

01:28:10.820 --> 01:28:12.519

Didong Li: it's not protected.

881

01:28:12.610 --> 01:28:23.979

Didong Li: But you can use some open source. The llama, which is open completely open and you can keep all your data locally. You can use those models for your like privacy related research.

882

01:28:24.240 --> 01:28:31.189

Didong Li: So at the end I would look, I'd like to thank chatty media itself, which is a main source where I generate the data

883

01:28:31.870 --> 01:28:34.299

Didong Li: of a coat and a slice.

884

01:28:34.340 --> 01:28:41.719

Didong Li: And up to reference, one is the data paper we used for the summary and the AI report by Stanford.

885

01:28:41.780 --> 01:28:44.119

Didong Li: So that's all for the discussion. Thank you.

886

01:28:50.630 --> 01:29:00.570

Didong Li: We're going to take a 5Â min break. We're going to allow our panelists to come up to the program, get settled, and then I'll be moderating the panel for your discussion.