   
Carlo Lipizzi: So it's 6, 32.

0:00

 Carlo Lipizzi: So I changed the the plan because some of you were not sure if if they could

0:04

 making time.

0:13

 the the courts were scheduled to end the this week.

0:16

 Carlo Lipizzi: but the very end at at Stevens, the final p of the we last until the beginning of may. So I thought that. Why don't we use the extra time to do things

0:22

 the best we can, so and that's why i'm giving you an additional week.

0:39

 Carlo Lipizzi: So I hope that you like the idea. But again, if for any reason, you want to close it to today, that's absolutely fine, I mean that that if you have your presentation, feel free to present it, that.

0:45

 Carlo Lipizzi: and otherwise we'll be next week.

1:01

 Carlo Lipizzi: So what do you think

1:05

 Carlo Lipizzi: any one of you time? What do you think?

1:14

 Kyle Jonas: Yeah. So I mean, I thought I just assumed the presentation was next week, because that's when the last module is.

1:20

 Carlo Lipizzi: So I was not ready for this week. So yeah, I was happy to see your your message. Okay, Good. Good, good, good

1:26

 Carlo Lipizzi: Christina. You okay?

1:33

 Christina Guida: Okay, Sorry. I'm. Actually, between 2 computers the computer I normally use.

1:40

 Christina Guida: We started on me and I jumped on my back up.

1:44

 Christina Guida: Okay.

1:48

 Christina Guida: I actually was under the impression that the presentation was done today. It was due today, so I was

1:52

 Christina Guida: kind of put in pedal the metal to get everything ready with my group for today.

1:58

 Carlo Lipizzi: Okay, we we can go today if you don't. If you don't mind, we're we're ready. No, no, no, no, no, I mean a. As in my email, if you prefer to do it today. Just make sure that you submitted it. So we don't have a

2:02

 Kevin Zeng: spending issue, and and I just want to be fully transparent with you a professor and i'm not sure it. It doesn't matter who who's who's a fall? It is, or is it? It is a fall? But on Sunday night. I think I saw some instructions posted on Module 13 for how to do the analysis.

2:19

 Kevin Zeng: but regardless, I think I think we're fine.

2:40

 Kevin Zeng: I just wanted to plan out there that we didn't realize that there was a you know, additional instructions for module and module 13. But we we we did started early.

2:43

 Kevin Zeng: I just want to to set that straight. Okay, then that's why we're ready. Okay, okay, go ahead. Alright. So I i'll. I'll share my screen. Thank you

2:54

 Carlo Lipizzi: sure?

3:03

 Kevin Zeng: All right, let's see.

3:04

 Kevin Zeng: All right.

3:09

 Kevin Zeng: We have a live audience today. You see my screen.

3:11

 Kevin Zeng: Yep. Okay. So

3:15

 Kevin Zeng: what what we did Our analysis on was on aviation. Specifically accidents that occur in the aviation space. I found that really, really, you know, our group found it really interesting.

3:18

 Kevin Zeng: The specifically it was a data set where we were able to apply filters to in a specific website, which i'll get into details later.

3:30

 Kevin Zeng: But we were able to partition. You know the the data set into 3 3 separate decades, and then we performed our analysis a month amongst those 3 decades. Right? So our our team is is Elise, Christina, and and myself, Kevin.

3:40

 Kevin Zeng: So this is a summary of what we did right. So the purpose is. Again, we were interested in the aviation accents over time to see how the accent that changed with a with a heavy focus on on injuries.

3:54

 Kevin Zeng: We decide it's best to analyze the data in in the 3 separate decades, right? Because it's a large data set right. And I I think it's started out in like 1960, S. And I. It Eventually the data for some reason stopped recording that around 2,015. But

4:05

 Kevin Zeng: our our our analysis was focused between like 19 eighties and 2,011.

4:23

 Kevin Zeng: So what we did first is a data preparation the aviation data contain. They have full inconsistencies. We we clean the data up before performing the analysis.

4:29

 Kevin Zeng: Obviously for the columns of dial that we were interested in. If it had missing data, we would just remove the entire role If there were columns that we're not print, you know, that did not apply to our analysis. We just left it alone, because it provide no value in just removing that.

4:39

 Kevin Zeng: So what we did. In Our strategy is to partition it, as you know, discussed into 3 separate decades. Right? So we have a one spanning 1979 to 1989,

4:57

 Kevin Zeng: 1990 to 2,000, 2,000and one to 2,011. So the the reason why we partition this data is so that we can perform our analysis and and see the trends right. The accent trends. My team will discuss that later on on on their approach to do that.

5:08

 Kevin Zeng: What we did last was we generate that right? We we pl at the data and comparisons between the 3 separate decades that we analyze.

5:26

 Kevin Zeng: and some methods we use identify columns that should be grouped together and identified, which comes that we can compare against.

5:36

 Kevin Zeng: So the that we we obtain data from Ntsb aviation query

5:44

 Kevin Zeng: we we left. We intentionally left the query right country, anywhere, state anywhere, month, all, and including all accidents, right, and the injury, severity would be fatal

5:50

 Kevin Zeng: that details. So in this data set we we have a a a good amount of columns, right? But we we focused on ones that made more sense for analysis. Specifically, events, data, engine type.

6:03

 Kevin Zeng: a far description which is the Federal aviation regulation description if it is present that means that that particular aviation rule is being scrutinized by this. That organization. Right? Let's say you had an accent, and it was related to maybe maintenance or something. Then there will be a a a. You know, a respective description to to indicate that

6:16

 Kevin Zeng: total fatal injuries, serious injuries, minor injuries, the phase of flight and the weather conditions which is pretty interesting. We decide to ignore certain ones because they're very unique per accident, like event. Id investigation, type, accent, number.

6:39

 Kevin Zeng: location country. Most of them fall under, like the United States. So it didn't really make sense to analyze the others, because there's only a few that that from from other countries, for for whatever reason, I guess it's more flexible out of the United States

6:57

 Kevin Zeng: that are preparation. So this is just like a brief summary of our strategy, right? We we did the filtering, and this our strategy for getting our 3 partitions

7:10

 Kevin Zeng: all right. We we had like a a data frame for a start, date and an end date right between those 3 3 years, right? Those 3 years start and start and and start end. And this is how we stored and returned our new data frame

7:21

 Kevin Zeng: for those those timeframes. Right? Of course, we ran to some issues where the data, the the date time fields were either empty or had a tab space. So we we had a strategy. We had a function to to filter that out.

7:37

 Kevin Zeng: That's it for me. I'm gonna pass this to

7:53

 Kevin Zeng: Christina. She'll talk about the the scope of our analysis now

7:59

 all right. So the

8:04

 Christina Guida: broad trends that we looked into were the types of injuries over the decades. the phase of flight versus the number of accidents.

8:07

 Christina Guida: the weather conditions, and how they affected the accidents, the purpose of the flight. There's a variety of different reasons why planes are flying, so what are they doing, and

8:13

 Christina Guida: how that affected the trends for accidents? The engine type for the airplane itself.

8:24

 Christina Guida: And then just generic accident trends over the years. Kevin, you can go the next slide.

8:29

 Christina Guida: Okay. So this is just the data that we looked at. This is just comparing fatal, serious and minor injuries over the decades.

8:38

 Christina Guida: The overall trend that we noticed at the highest number tended to be in the 1990 to 2,000 category, a decade that held true for total overall injuries, minor injuries, and serious injuries.

8:47

 Christina Guida: The only one that was a little different was fatal. Injuries actually had a higher

9:00

 Christina Guida: number in the 1,979 to 1,989 decade

9:05

 Christina Guida: for all categories. The lowest decade was 2,001 to 2,011,

9:08

 Christina Guida: and across each of the 3 decades as well. The highest injury type was fatal.

9:14

 Christina Guida: You know the next I have.

9:23

 Christina Guida: Alright, so this is just the correlation analysis that we ran on the data for each of the decades. So it's actually comparing

9:26

 Christina Guida: the types of injuries to each other, and seeing the relationship between one type of injury and another in both 1979 to 1989, and 1990 to 2,000. There was all positive correlations between the types of injuries.

9:33

 Christina Guida: So, as one injury type increase, you'd expect to see an increase in the other injury types as well.

9:49

 Christina Guida: The strongest correlation in both sets of both decades was between serious injuries and minor injuries. 2,001 to 2,011 was a little different. There was still a very small positive correlation between serious injuries and minor injuries.

9:55

 Christina Guida: but the other combinations actually saw negative correlations between them, so, as one increase, the other would be expected to decrease, which was

10:12

 Christina Guida: vastly different than the previous 2 decades.

10:20

 Christina Guida: But the overall turn was actually a weakening of the correlation over time between the injury types.

10:24

 Christina Guida: All right, Kevin, go on the next one.

10:31

 Christina Guida: So in this chart we're just comparing the number of accidents that occurred at each phase of flight.

10:34

 Christina Guida: For each of the 3 decades

10:41

 Christina Guida: they look kind of similar across the years, the top phases of flight for injury.

10:43

 Christina Guida: for accidents, i'm sorry, would be cruise, maneuvering and take off.

10:49

 Christina Guida: and the lowest we're seeing the lowest number of accidents are seen echo around, landing, standing and taxing across all 3 decades.

10:54

 Christina Guida: All right, Kev.

11:04

 Christina Guida: Thank you. Thank you. This is I got one more at least Takes over an 11 last one. So for weather conditions we also compared them to face the flight.

11:05

 Christina Guida: So across all 3 decades we saw a trend where the highest number of accidents actually occurred in good weather, conditions visible.

11:20

 Christina Guida: visible, meteorological. That's a tough word conditions. So when the pilot's actually able to rely on what he can see as opposed to having to rely on the instruments.

11:29

 Christina Guida: and at first we kind of thought that was a bit alarming. Is it something up with the pilots? But it actually may be a factor of the fact that statistically, most flights actually occur in good weather a lot of times. If there's

11:40

 Christina Guida: poor weather or visibility flight, so we cancel delay. They're not going to take off. So it may just be a factor of the number of flights that actually occur, some statistically be more likely to have an accident there

11:53

 Christina Guida: that that was actually something. We said that with another data set might be an interesting comparison to look into.

12:05

 Christina Guida: But across all the decades as well, we noticed that in maneuvering was the highest number of accidents that occurred even during good weather.

12:12

 Christina Guida: All right. i'm going to pass it over to at least. Now

12:22

 Kevin Zeng: Yeah, go for at least

12:25

 Elyse Spinelli: Alright, perfect thanks, guys. So now we're going into specifically over each decade, the purpose of flight.

12:26

 Elyse Spinelli: So we see here there is a trending very large percentage

12:32

 Elyse Spinelli: which is represented by the purpose of personal flying. So from 1979 through 2,011 this is represented by approximately 50 to greater percentage from that 1,979

12:36

 Elyse Spinelli: to 2011 window. Now.

12:52

 Elyse Spinelli: being a factor of you know, thinking what is the most common purpose of flight

12:55

 Elyse Spinelli: about the time of the sixties is when commercial flights started to become popular, and so gradually air travel was less exclusive. By about the eighties via a popularized, You know, commercial flying experience. And so into the 2 thousands more people had access

13:00

 Elyse Spinelli: to plane travel, had the money to also had via the Internet. You know more of a desire to see places that were farther away.

13:17

 Elyse Spinelli: So that is

13:26

 Elyse Spinelli: a reasoning for that. 50% approximate growing to so much larger to about 70% over the years. And then the subsequent larger percentages here are represented by in order, business, unknown purpose, and instructional flights.

13:27

 Elyse Spinelli: and so obviously business can be justified by saying that this is a necessary means of flight in terms of obviously those who are employing that need to travel for their.

13:46

 Elyse Spinelli: for their work

13:58

 Elyse Spinelli: would quantify a mass amount of people. Then we have an unknown reasoning which could be for any very smaller and significant purpose of flying, and then instructional flights. For obviously there are

14:00

 Elyse Spinelli: pilots that are in school that would be flying planes where

14:12

 Elyse Spinelli: any number of situations could possibly go awry.

14:16

 Elyse Spinelli: So the next slide, please, Kevin. So now we move into the engine type of what the plane would have. So here

14:20

 Elyse Spinelli: all decades, analyze the engine type with the most injuries would be the reciprocating engine. Aka, a piston engine which uses one or more pistons, you know, convert pressure into a rotational motion and a rotational source of power.

14:28

 Elyse Spinelli: So

14:43

 Elyse Spinelli: over the years we see that this actually grows the use of this engine, and it is

14:44

 Elyse Spinelli: that the piston engine has become the most popular type of engine to date. So that is a very literal reasoning for

14:52

 Elyse Spinelli: this percentage going up. And so

15:01

 Elyse Spinelli: also the reason for this going up is that, you know, advancements in technology have happened over the years which supports us, the use of this type of engine

15:05

 Elyse Spinelli: more so over any other type.

15:16

 Elyse Spinelli: so you can go to the next slide. So now we move into.

15:19

 Elyse Spinelli: They look in the Federal aviation regulations for the far rules. So these are sets of rules set by the Federal administration to

15:23

 Elyse Spinelli: kind of look at this data to see how every regulation was documented over each decade, and which rule was under the most scrutiny in terms of looking at accidents that were reported.

15:34

 Elyse Spinelli: So we did notice a majority of accidents that were looked at. We're missing a data value, regardless of what year we were looking at.

15:48

 Elyse Spinelli: and also the rule that was under the most scrutiny over all decades by the Federal Aviation

15:56

 Elyse Spinelli: Association was general aviation. Aka. These personal flying

16:04

 Elyse Spinelli: instances

16:09

 Elyse Spinelli: Some next slide.

16:13

 Elyse Spinelli: Oh, you're done. At least we're going to do a quick conclusion just to wrap up. So, for over all 3 decades the greatest number of accidents that we saw occurring was during that Vmc. Period, or that visual meter meteor lodge. Oh, my goodness, Christine, that you're absolutely correct

16:15

 Elyse Spinelli: conditions. So, just to reiterate. That was when the pilot was able to use good visual cues to be able to navigate the aircraft.

16:34

 Elyse Spinelli: and

16:44

 Elyse Spinelli: therefore would have you know, a better sense of surroundings when in flight. Another trend that happened to occur for all 3 decades was that the most occurring injury type was that of fatal.

16:45

 Elyse Spinelli: And for the if they are data values, President

16:57

 Elyse Spinelli: present the general aviation category which could mostly categorize that personal and business purposes of flying was scrutinized the most, and was therefore put under the most investigation.

17:01

 Elyse Spinelli: And now we'll open anything up to more questions from everyone.

17:16

 Carlo Lipizzi: Okay, great. Thank you.

17:20

 Carlo Lipizzi: So I have a few questions. So one is more a general question. Why, you didn't consider the making model.

17:23

 Kevin Zeng: Oh, yeah, so I we we explain. In the report.

17:35

 Kevin Zeng: There were a lot of nick and models to the point where it wasn't. It was very difficult to to consider. I mean, I guess it can, but it would take a lot more effort, not sure what value would bring.

17:38

 Kevin Zeng: but from from what we we try doing. They They end up having a lot of

17:50

 Kevin Zeng: variations making models in the data set.

17:55

 Elyse Spinelli: And I guess if you were going to move into the type of plane. A piece of that aspect that we took was being the engine there like being that there were a much more succinct data set to look through, and you were able to get numbers that put more value toward the analysis.

17:59

 Carlo Lipizzi: Yeah, it's kind of interesting because i'm a former student of mine did the in the 800 project on a a a similar problem. She

18:17

 Carlo Lipizzi: he's working, I think, for lock it.

18:30

 Carlo Lipizzi: and and then she did the the on the same topic. So i'm not saying that they became an expert. But I a little bit familiar with the

18:33

 Carlo Lipizzi: the the topic on your project. The second question is why you think there are so many less fatal accidents in the last decade

18:50

 Kevin Zeng: I can go first. If

19:08

 Kevin Zeng: yeah, I can follow up I

19:11

 Kevin Zeng: the investments in technology, right? Maybe the robustness and the design of like the landing gear. For example, I I I've seen a lot of documentaries

19:16

 Kevin Zeng: because I just like it on documentary nerd, but they they just made everything better. There's more strict regulations, more may, maybe more maintenance, right the panel. There's I have a cousin that works in the maintenance side of of aircraft, and I think the requirements are more stricter, right? Just everything overall.

19:24

 Kevin Zeng: It's just a better over the years.

19:43

 Carlo Lipizzi: Yeah. I mean that it it's a quite a relevant point, because when you do with an analysis like what you did, the the the main question is why you are doing it. Who can benefit from it?

19:46

 Carlo Lipizzi: So when you see those differences in breeding down on why there are a such a difference. So it could be an indication, I mean, if it is a a matter of a a different regulatory system, what are the regulations that that impacted the the most.

20:03

 Carlo Lipizzi: and that the second answer, I think kind of a following the first one. How can I reduce even in farther? So did did they reach a a sort of plateau for those changes.

20:26

 Carlo Lipizzi: So I mean, i'm not at all like criticizing the very good work that that that you did but keep in mind that that you can get some actionable. So somehow, from a research like like this one.

20:41

 Carlo Lipizzi: and the actionable in this case would be okay. I see a reduction. Let let me better understand why

21:00

 Carlo Lipizzi: a a and a. Is there any margin for improvement? So those things could be the next step for the analysis that that that you did not not saying that that is not great, but it it's really good.

21:09

 Carlo Lipizzi: The other question I have a, and that's unfortunately something that is a common in mostly analysis, that that I see the lack of a normalization by the size of the population that you are comparing.

21:24

 Carlo Lipizzi: So if you don't normalize for a a number of flights, so

21:41

 Carlo Lipizzi: then the comparison may not be that much relevant. So either you go in in terms of percentage, or you go with the normalization with with the that there is the sort of a person that January.

21:48

 Carlo Lipizzi: So what do you? Because at the certain point you mentioned also the the fact that the numbers can be different, because the number of flights in the different periods could be different. So how did you consider that

22:06

 Kevin Zeng: I'll be honest, Professor? We do not consider that, but that's a very good point. I I I think, what you're saying, it makes absolute sense. We should probably

22:26

 Kevin Zeng: do, maybe like an approach where we do 1,000 from this year, 1,000 for this decade, another 1,000 from the second decade, another 1,000 from the third decade for some type of of a way to make it equal comparison.

22:34

 Carlo Lipizzi: Yeah. I mean that in any form of a a normalization would would be fine. Because I I mean, if you had the percentage that the percentage, in a sense, is the normalization, because you have a the the the the type of

22:48

 Carlo Lipizzi: A. As a percentage of the total. What they are, they total is.

23:05

 Carlo Lipizzi: and then you can compare them

23:11

 Carlo Lipizzi: order. You can normalize using a a any form of a normalization, the meeting, Max, or whatever is the most, the the

23:13

 Carlo Lipizzi: the normalization that that that you pick. But I mean, like, keep in mind that that that's something that you may want to have a a for future analysis, because otherwise, if they feel to to compare.

23:25

 Kevin Zeng: Thank you. That's very good feedback. I appreciate that.

23:39

 Elyse Spinelli: To go back to your second question also about you know why you might have the injuries

23:42

 Elyse Spinelli: and accidents that, like went down over the years, I would say to there's definitely, besides the less human factors of new technologies. There is also probably better training and risk mitigation factors with

23:48

 Elyse Spinelli: instructing new pilots and teaching like new pilots, how to flying. Also that I imagine

24:03

 Elyse Spinelli: there is a more of a demand for pilots, and so, therefore over greater number of them there is a a safer flight.

24:09

 Carlo Lipizzi: Yeah, in the work we did with the with the other student. We know this the same fact that that accident happened more in clear of sky.

24:18

 Carlo Lipizzi: and it it was a kind of surprising, so

24:30

 Carlo Lipizzi: she didn't do much of the normalization that would help, but because I mean, if you have all the data that are in clear of sky and few that are with cloud this guy, then obviously you will have more on the clear sky cases.

24:38

 Carlo Lipizzi: But another consideration that we did was. and when you go in clear sky condition. You rely on the equipment.

24:55

 Carlo Lipizzi: and the the pilot may pay less attention because there is nothing to be worried about.

25:11

 Elyse Spinelli: and that could be a another factor. Yeah, I I went to note that that's exactly what I me and my team were discussing specifically. Christina, I was saying. We can add that to the conclusion how they could be less reliant on technology. But maybe they should use a combination of the technology and the clear skies to make a judgment.

25:20

 Kevin Zeng: So that's another good point.

25:40

 Carlo Lipizzi: Yep. Yep. Okay, sounds great. You did good. You'd be very good.

25:43

 Kevin Zeng: How big was your code, then? Do you have the code? The By the way, yeah, yeah, we have the code. I I believe, minus the the comments. The code should be around 350 lines, but not

25:50

 Carlo Lipizzi: Okay? So as a final conclusion, was this courts useful for you?

26:04

 Kevin Zeng: Yeah, absolutely. I think it. It forced me to learn python, but certainly it. It's a new. It's a new language for me. And do we learning the different? You know, various ways to process data and the what? What was certainly interesting is the the analysis that we had to do.

26:13

 Kevin Zeng: I I never had to do analysis in in this kind of way before, so it definitely was a a new, I opening experience for me, and and of course allowing us to choose a topic of our choice, for the final was very enjoyable experience.

26:31

 Kevin Zeng: That's for me

26:46

 Carlo Lipizzi: good Christina Elise.

26:47

 Elyse Spinelli: you like I before. They only used Python a

26:50

 Elyse Spinelli: a handful of times for very specific, like AI purposes.

26:54

 Elyse Spinelli: but because they were so specific. I I kind of lost touch with it quite quickly after ending the assignment. So

26:58

 Elyse Spinelli: I certainly feel like I actually put, you know, methods to memory, which is for me pretty impressive. I only usually code with like Bash scripting at work. So

27:05

 Elyse Spinelli: I think this, you know everything culminating together in this project

27:14

 Elyse Spinelli: was very useful and made sense to You know how we moved through everything.

27:18

 Carlo Lipizzi: Good. Good.

27:25

 Christina Guida: Yeah. I don't use. I probably haven't looked at coding since my underground number of years ago. So it's definitely

27:27

 Christina Guida: different than my norm. I was

27:36

 Christina Guida: interesting to see different ways of analyzing data. I'm generally in excel person. So it was interesting to see I could find the information in excel, but I could also do it another way for other purposes.

27:40

 Okay.

27:53

 Carlo Lipizzi: okay. So I I mean I I was not fishing for compliments. It was just getting your input because I mean this: this courts is running, since quite a while

27:56

 Carlo Lipizzi: I

28:09

 Carlo Lipizzi: change a little bit each semester and getting input. It's really important to me. So is there any part that you think I could could have been different or didn't work Well, as you expected, the

28:11

 Elyse Spinelli: not

28:31

 really

28:32

 Elyse Spinelli: not necessarily

28:33

 Elyse Spinelli: my head right now. Not necessarily.

28:36

 Kevin Zeng: Yeah, not not for me. But I I was actually curious. Oh, Professor, is there any class that incorporates

28:41

 Kevin Zeng: students to contribute to like a code repository like a collective, you know way of developing or working on something or an assignment.

28:47

 Carlo Lipizzi: not really. We have more research projects right now. As an example, I'm. Working with some students on a project to create

28:58

 in a Ssc. Chat box based on on a larger language model that could be sort of a a Tudor for a students.

29:13

 Carlo Lipizzi: and we are going to start with the with this course with the 6, 24, and and then we use the from the classes as data along with the the the teaching material so the slides, the readings and all the rest.

29:26

 Carlo Lipizzi: All of that. We'll generate code, and they code will be available to students as well. We don't have a real repository.

29:49

 Carlo Lipizzi: I mean that

30:02

 Carlo Lipizzi: repository for the the the

30:03

 Carlo Lipizzi: wouldn't

30:08

 make not much sense. I share a

30:10

 Carlo Lipizzi: some parts of code during the the like. I don't know what cloud or cleaning data, so that those are either scripts or as nippets. It's something that they share.

30:14

 Carlo Lipizzi: Probably it would be good that to create a sort of small a gee tab with the all those pieces. We, our students can just go there and get it

30:34

 Carlo Lipizzi: for the larger scripts or problems. I mean.

30:46

 Carlo Lipizzi: they go into research projects, and for those we do have a repository. But then the the research projects sometimes have some some restrictions.

30:55

 Carlo Lipizzi: We do research properties for the Dod, and a while at the results are the public, the the code, the maybe now after. So that's

31:12

 Carlo Lipizzi: my experience. So far.

31:26

 Kevin Zeng: all right. So it's really interesting that you mentioned the the chat boss because i'm working on a a software as a service platform. And we're also trying to incorporate. You know a AI into into that as well. More and more of so to

31:33

 Kevin Zeng: facilitate the whole, it help desk that we're shooting, you know, basically automating the

31:48

 Kevin Zeng: the the way we we handle requests some request to be automated to a certain degree.

31:55

 Carlo Lipizzi: Yeah. I mean that this paradigm of the larger language models it. It's very interesting because

32:01

 Carlo Lipizzi: you can have an interaction with the data that is more conversational, that is, lowering the threshold for people to use it.

32:12

 Carlo Lipizzi: We are a considering. I mean

32:22

 Carlo Lipizzi: the model that they developed on on a specific

32:30

 Carlo Lipizzi: knowledge base that is a a specific one domain as the limitation of not having the the common sense. So I think so. That seems to be obvious, may not be noticed by the system, and that it's why we want to incorporate the the existing

32:42

 Carlo Lipizzi: large language models either open AI like or a those from a hugging face, and they have a some models that are available

33:07

 Carlo Lipizzi: for the common sense meaning. Before going into the analysis that is more a domain specific, we go to the common sense to eliminate the obvious.

33:19

 Carlo Lipizzi: It would be great to have a resources enough to have something that is a domain specific and big enough like bloom. But the so Bloomberg had the the

33:32

 actually all the I don't know how many years, all financial

33:48

 Carlo Lipizzi: data and publications, and they use that to create the the entire model.

33:55

 Carlo Lipizzi: We don't have that. So even with the all the transcripts and all the the material. It would be a tiny, tiny fraction compared to what could give the the system a a real, enjoyable interaction.

34:03

 Kevin Zeng: Yeah, it's as good as the that you have right.

34:21

 Carlo Lipizzi: It is a separate project is to use. I mean, the large language models have a component that is generating a language.

34:29

 Carlo Lipizzi: The idea is, to use the language, generation, capability of a large language model for presenting the results of a different model.

34:44

 Carlo Lipizzi: So I developed my mode a lot with the let's say, a combination, let's say, generating the data like those that would go in a dashboard.

34:56

 Carlo Lipizzi: So I have quite a lot of data, and I could create a a a dashboard instead of creating the dashboard. I feed the a language model, and the model will present the results in plain English.

35:09

 Carlo Lipizzi: and that could be useful in a when you are in critical conditions. So you are on the bottom field. You are in a intensive care unit.

35:23

 Carlo Lipizzi: You are a in a nuclear plant with an emergency. So in those cases you you don't really rely much on what you see you are doing something, and you want your explanation in English.

35:35

 Carlo Lipizzi: So when you have those models so generating the text, then the text to speech is a no-brainer. There are a a medium module for that.

35:54

 Carlo Lipizzi: So that's another thing that that we are exploring there so language as a an equivalent to the visualization of the next a blow, something like that.

36:05

 Kevin Zeng: Yeah, Professor, I have 2 questions. Thanks for asking

36:15

 Kevin Zeng: some some of my work pose a question about language models. Specifically, I think they're concerned about how it would drive. You know it. And and is that something we're we're we're you know, concerned about, or adapting right where, where your thoughts on on that is, that is that the future is that going to define?

36:18

 Carlo Lipizzi: Well, I I mean, we. We need to set the expectations in the proper way.

36:36

 Carlo Lipizzi: We tend to have a way too much expectation from technology. So when the first computers we're used there, we are people saying the machine side that and has to be through.

36:45

 Carlo Lipizzi: But the reality, the algorithm that the machine was using and the data that the machine was using. You know we are not so

36:58

 Carlo Lipizzi: well done, or a strong or tested or larger to, I mean, justify the fate.

37:07

 Carlo Lipizzi: We are pretty much in the same situation here. So those models

37:15

 they call them

37:23

 Carlo Lipizzi: generative models, but they're not generating anything they are generating. Answers to that are the equivalent of what Google is giving us

37:24

 Carlo Lipizzi: with the layer of conversation.

37:36

 Carlo Lipizzi: So this is to be very intelligent, but in reality they're not

37:39

 Carlo Lipizzi: erez agmoni, creating something like an intelligent mind that could do. They are as teaching together pieces, and presented it 150 in a sort of a a probabilistic way.

37:44

 Carlo Lipizzi: If we continue increasing the size of all the models we will be. I mean, we we will have a both, so that are more accurate in analyzing the date than presenting the data.

37:59

 Carlo Lipizzi: But there is no generation of a new concept, new ideas.

38:18

 and there is no domain.

38:25

 Carlo Lipizzi: So we didn't solve the the the problem. All the the specific knowledge that you can have in that one specific domain.

38:28

 Carlo Lipizzi: The

38:40

 Kevin Zeng: yeah, If you have it, it should be relying on the the the basis of

38:41

 Kevin Zeng: of what it it it knows right to to generate the answers. Yeah, I mean, if you add the all your knowledge in a

38:46

 Carlo Lipizzi: the

38:58

 Carlo Lipizzi: aircraft industry, and you dump into an open AI.

39:01

 Carlo Lipizzi: It will be a drop in the ocean, meaning the part that it will discover will be sort of a deluded in a homeopathic way in a way that there will be no trace of the specific knowledge.

39:13

 Carlo Lipizzi: So what we need to do is to come out with a better

39:32

 Carlo Lipizzi: or more representation of the knowledge that can value the different components, so the generic and the specific in the proper way.

39:39

 Carlo Lipizzi: And then I mean that we will be able to reach the the point that we will have those types of representation working in a in a a proper way, most likely. Yes, probably not in the next couple of years, but in in the next 10 years, probably.

39:51

 Carlo Lipizzi: or even 5 years. At that point it would be interesting, because I mean.

40:14

 Carlo Lipizzi: I don't know if I already mentioned that i'm writing a book on a societal implication of a. I am machine learning.

40:25

 Carlo Lipizzi: and I am analyzing those issues.

40:34

 Carlo Lipizzi: jobs will be destroyed, but jobs will be created. So if you look back

40:40

 Carlo Lipizzi: to to the past revolutions driven by technology.

40:47

 Carlo Lipizzi: the revolution, so they did destroy the the entire categories of of jobs.

40:53

 Carlo Lipizzi: If you think the the music industry who is buying buying it. Yes, there is a little bit, but it the marginal business. So who is buying a cassettes? Who is buying a analogue? Music in general.

41:05

 Carlo Lipizzi: blockbuster renting Dvds or Cassettes that they are gone or developing teachers printing pictures. So all of those were big businesses, but they are gone

41:23

 Carlo Lipizzi: even now with the streaming. I mean videos move these.

41:42

 Carlo Lipizzi: They they they! They they are gone, but new industries will come. So we really need to be very flexible and very open

41:50

 Carlo Lipizzi: and fast reacting, because things will change. Things are changing very fast. The major issue that they really see is more

42:01

 Carlo Lipizzi: on the concentration of

42:13

 that that we see more and more.

42:17

 Carlo Lipizzi: If we stay with this paradigm like open AI. There will be 4 or 5 companies able to create those giant models.

42:20

 Carlo Lipizzi: and we will need to use them. Yeah.

42:35

 and that's an oligopoly.

42:39

 Carlo Lipizzi: but it would be a noticeably of the knowledge, because we will all use the know that that is collected by them.

42:43

 Carlo Lipizzi: So with the chat gpt, the vast majority of the data is in English. That that means that we have a a, a, a large section of the population in the world that is under represented.

42:51

 Carlo Lipizzi: So we don't want that

43:13

 Carlo Lipizzi: there are many changes that we need to do. The current proof of concept that we are experiencing with with with the open AI, and similar

43:15

 to to make it work.

43:28

 We need to be aware of what what is wrong, and try to do our best to work better.

43:32

 Carlo Lipizzi: So distributing intelligence is something that is coming up

43:39

 the

43:45

 I mean.

43:48

 Carlo Lipizzi: In fact, we already have it with the I mean most of our phones. They have a a dedicated components for our machine learning meaning, and there is some computing 150

43:51

 Carlo Lipizzi: on the machine learning side that is happening in our phone, and probably on our smart watches, 250. But we need more than that. So we need more of a distributed computing, more of a distributed intelligence.

44:06

 Kevin Zeng: So anyway, that's going to be an interesting period, though. Yeah, we're living an interesting times.

44:24

 Kevin Zeng: Yup, yup, Yup, yeah. I'm always interested on how how the you know. Does the the data preparation that cleaning for for their learning models? Very, very interesting stuff. My, My, my last question is so nice exciting is I? I message you and the Ta. About submitting the file project. We just need one person to submit. Everything. Is that correct?

44:32

 Carlo Lipizzi: Yes, yes, Yes, yes, so I mean that

44:53

 Carlo Lipizzi: just just to be sure that they everything will be done in the proper way.

44:59

 Either one of you would submit, or all of you would submit the same thing.

45:06

 Kevin Zeng: Okay, the last. The last time I checked the submission button was not available. But you know it could be available. Now. Yeah, okay, All right, Thank you.

45:13

 Carlo Lipizzi: Sure. I mean the the reason why I would be more

45:23

 Carlo Lipizzi: on all of you with submit is because if at the certain point there will be an audit. I don't want someone that checking and a normally. Why, those students didn't submit.

45:30

 Carlo Lipizzi: so I mean we know it, but it could look

45:45

 Carlo Lipizzi: better if you all will submit.

45:53

 Kevin Zeng: Okay.

45:58

 Carlo Lipizzi: And it would make a much professor that you that that you who will will will get it.

46:01

 Kevin Zeng: Thank you, Professor, for sharing all your insights

46:08

 Carlo Lipizzi: absolutely, absolutely. So again. Next week there will be other presentations, so feel free to join my opinion. It's a good learning moment, because you would see what I did the on probably

46:10

 different areas, different topics.

46:30

 Carlo Lipizzi: So I

46:32

 Carlo Lipizzi: encourage you to join us next week.

46:34

 Kevin Zeng: Okay, Thank you for us, sir.

46:39

 Carlo Lipizzi: Okay, thank you all.

46:40

 Carlo Lipizzi: If you don't have a question. So that's the end of this sort of a class. This is a very short one.

46:44

 Carlo Lipizzi: and I appreciated your presentation. You did a a great job.

46:52

