NCIIPC AIGC - Webinar on Problem Statement No. 3 at 11 am on 25th Aug. 2025 (Monday)-20250827_061054UTC-Meeting Recording

August 27, 2025, 5:21AM 48m 59s



PO PS - 07 0:50

One language.

Arabic, Chinese, English, French, Russian, Spanish.

Arabic Chinese.

English.

French.

Russian.

Spanish.



Adithya Umare 4:41

H5 and F5.

I.

Yeah.



RM Raj Mehta 9:45

Good morning everyone. I'm Raj Mehta from FITT. I hope that I'm audible. We should start the meeting soon. Good morning. Since you're since you're getting members of like applicants continuously in the.



Jyotimaya Mohanty 9:52

Hi, good morning. Yeah, yeah, I heard about that.

RM Raj Mehta 10:01

Maybe not, so we should start the meetings shortly.



Jyotimaya Mohanty 10:06

Yes, yes, yes.

Bhavya Patel 10:58 Oh.

Mhm.

Meenakshi 11:27

Yeah.

Cortana.

RM Raj Mehta 17:02

OK, let's start the webinar. I welcome you again. I'm Raj Mehta from FITT. The webinar go in the in the flow that first the officer in charge, Dr. Ashtalakshmi shall go through the problems.

Problem statement 7 document. She will tell you about the background and the different requirements and the different types of datasets, the dates of release and what are what are the expectations.

From the applicants regarding this particular PS.

So I give it over to Doctor Srashmi.

That's Asilakshmi over to you.

PO PS - 07 17:50

Yeah. Good morning, everybody. Am I audible? Yes, yeah, yeah. So this is a problem statement 07. It's about.

- Raj Mehta 17:55 Yeah, you're doing.
- vani 17:55 Yes, yes.
- RM Raj Mehta 18:02

I request I request all the other members to kindly mute themselves.

All the other members in the audience currently mute themselves.

Thank you.

OK, so it is problem statement 07 about password recovery using AI and ML file. As you people are all aware of that despite multiple rise of this multi factor and biometric authentication, password still remains as a dominant thing in securing the. Files and conventional approaches like brute force, dictionary and rule based are all increasingly becoming inefficient even with high resource. So since this AI is changing the game, we thought there are some approaches also already available in open domain.

We have posed this problem, so our mission is to build an Al powered platform for smarter and faster efficient way to.

Way to outperform this traditional methods which are already available like dictionary attacks and other things to recover this passwords. OK and so now we are we are having at stage one we will be having the password protector file especially we are looking for.

Office 2013 and above versions and PDF 1.7 and above versions. So we are only concentrated on these type of files when it is password protected. So we are expecting you to use an AI and ML technique to recover the passwords of this password protected files.

So in stage one you you will be having the the time till September. So on 15th of September at around 10:00 AM we will be releasing some 50 files to you which is password protected for as a mock test.

And after that in stage one there are three levels in which in level one we will be releasing 25 set on 26th of October and for level 220 files again on 1st of November and Level 3.

When.

It finds again on 6th of November. So as soon as you have recovered the password, you just upload your password in the portal given to you because the timestamp is also will be recovered. How fast do you have recovered the password? So the weightage of the things will be like 75% for the number of passwords you have recovered and the time taken will be given as a 25% weightage. But in this you have to note mock is vector is not for.

Validating we will not evaluate your mock test. You have to work only on level one, level 2, Level 3. Why we are giving mock test is just for your hands on and practice and to see how you have performed or how your model is working. For that purpose only we are giving mock test so.

That will help you or you will understand because the leader board for that is also there so that you will understand how many people around you have done the problem and how many success rate was there. So this is the thing and at the end of the Level 3.

Based on the highest score, we will be selecting 15 to 20 candidates who will be called here to Delhi and they will be given the same platform because you people will be using different platforms for your implementation. So there we will be giving you 800 GPU cards with 40 or 80 GB and CPUs of 48 core or.

Something like that and the RAM with the minimum were 256 and above and the operating system will be Ubuntu. You have to implement your whoever selected the 15 to 20 candidates will be implementing their solution here and again we will give the test vectors here.

And based on that, six participants will be selected, so any queries?

RM Raj Mehta 22:08

Thank you, Doctor. I request members in the audience, whoever has any question, Corey, they can raise their hand in the webinar. Doctor Ashlakshmi shall take the questions 1 by 1.

So we have first hand by Mr. Soru. Yes, Sir. So, so, yeah, yeah.

Swaroop M N 22:29

Hello good morning. Whether the files that you're looking is could be a zip as well as encrypted or it is just a raw PDF and doc file.

PO PS - 07 22:43

Oh ma'am, for the first stage we are only looking into Office 2030 and PDF. We are not looking into zip.

SN Swaroop M N 22:53 OK. Any encryption over there?

PO PS - 07 22:53

So all exact, yes means whatever the file will follow, we will protect that file with some password that that file will be using the inbuilt encryption. Suppose if it is

Office 2013, Word will be using some encryption.

And every version uses their own encryption method, like AES 1262128 or 256, whatever is the standard for that file.

- Swaroop M N 23:25 Thank you.
- Raj Mehta 23:26

 We have next Anupam. Yeah, yeah.

Anupama Itikirala 23:28

Ah, Anupama here.

Anupama myself.

My question is like how should we submit this solution?

Is it an Excel sheet only the results or should we submit the source code that we have produced on that?

PO PS - 07 23:46 M.

Sorry, can you please repeat?

Anupama Itikirala 23:54

My question is how how it should be some because is it a container application? How should we submit our solution?

PO PS - 07 24:01

Yes, yes. Like once you have to submit the full document set first at the first point, as soon as your password is recovered, you have to submit your password only. So once all the things has been done after after 10th of November or so.

You have to submit there. We are forwarding all the details how you have to submit your that will be shortly published. I think what are the things we need from you and especially your docketized model and the methodology used your model architecture, why you have selected the model, what is the novelty of your? Model, what is the scalability? We will be putting one questionnaire very soon in the site and you have to answer that and you have to submit your dockerized model.

Anupama Itikirala 24:44

When we say all these things, am I expected to provide the design methodology design document?

- PO PS 07 24:55 Yes.
- Anupama Itikirala 24:57

OK, the next question I have is can we use when you say artificial intelligence, there's a close match to the NLP. Can we use any of the NLP open sources like hugging force, hugging face to understand the model?

- PS 07 25:11

 Yes, you can. You can use any open source this thing, but at the final end stage you should it should be like you should be offline, yeah.
- Anupama Itikirala 25:16 Offline.

Yeah, because I think this I can actually embed it into my application container as it. So I I'm at any of the LL lamps to trying the data set for this.

PO PS - 07 25:27

Yeah.

Yeah, right now you can do it, but at the end of the stage it should be offline. If you can do without your online integration, then it is fine.

- Anupama Itikirala 25:40
 It's not online, just offline. I will download the version that I'm using and.
- Po PS 07 25:43
 You can use anything.
- Anupama Itikirala 25:46 Thank you.

- Raj Mehta 25:51
 Yeah, we have next, Mr. Aditya.
- AU Adithya Umare 25:55
 Uh, hello. Yes Sir, am I audible?
- Raj Mehta 25:58 Yeah, you're ready.
- Po PS 07 25:58 Yeah.
- Adithya Umare 26:00

 So my question is in PDF and office files, there are two types of password owner protected and user protected. So which type of passwords means would there be a mix of the files or all the files would be of a certain type of password like user protected password?
- Yeah, both we are looking into it. If somebody has protected both or somebody might have protected only with the owner. So if it is only protected with the owner, it is sufficient if you it has been protected with the user. See once you break the owner password you will get entered into the file.

 And after that user might have protected with the rights only like whether it is
 - readable only or write option or something editing option. So we are but we are looking for both. If it is protected by both, both should be recovered.
- Adithya Umare 26:50 OK, ma'am. Thank you.
- Po PS 07 27:03 Yeah.
- SWaroop M N 27:07

In annexure B in the document there is something called mock test vector submission and there is something called application ID. I did not understand what is this application ID.

PO PS - 07 27:14

Yes.

Application ID is your application. What when you have registered? If you have got some application ID so that we will recognise you who you are.

- Swaroop M N 27:42 Sorry, still I did not get the application ID.
- PS 07 27:47

 I think Raj Mehta, Sir, you can help them out. Like when they have registered, they might have received some ID so that we can identify who have submitted the problem statement from the startups, all the startups, there is some application ID Sir.
- Raj Mehta 28:03

 Yeah, there is an application ID which which which it starts with a IJC number.
- SN Swaroop M N 28:03 Yes, yes. Yeah, there is one. Yes. Yeah. Yeah. Got it there. There are one of the occupation.
- PO PS 07 28:03 Not.

Yes, that application ID so that we will know who has submitted the result for the answer.

- Swaroop M N 28:13 Excellent. Thank you.
- RM Raj Mehta 28:19

Had raised rutuza.

Rutuja.

Rutuja_aerovania 28:25

Yes, I had a question that for stage one will file strictly be in Office 2013 plus a year or say maybe PDF 1.7 plus a ES or should we also prepare for legacy crypto modes like RC core or older office formats?

PO PS - 07 28:43

No, no, only these 2 files we are looking into Office 2013 and PDF.

Rutuja_aerovania 28:48

There is a.

1.7 so we we shouldn't be prepared for RC 4 type of crypto mode Z.

PO PS - 07 28:52

And it is.

No, no, actually why? I told the a ES is I am not. I am not using an additional a ES on that whatever the file format. Every file formats follows some encryption like RC 4 is followed in old office, new Office, Office 13 and double so.

- Rutuja_aerovania 29:12 Right, right.
- **PO PS 07** 29:15

Whatever encryption the each file type is following, you have to look into that, but we are looking only into office 2030 and PDF 1.7 and above for the stage one.

- R Rutuja_aerovania 29:28 OK. Thank you.
- Lakshit Verma 123196938 MITMPL 29:39
 Hi I had a question.

- Raj Mehta 29:41 Yeah, please.
- Lakshit Verma 123196938 MITMPL 29:44

 Our participants allowed to use pre trained language models like GPT or must everything be trained from start?
- PS 07 29:54
 Sorry, can you please repeat?
- Lakshit Verma 123196938 MITMPL 29:57

 Can partitions use a pre trained models? Almost everything we trained from scratch.
- PS 07 30:04

 Pre trained model if it OK you can use it if it has been pre trained but at the end of the day you should. It should work.
- Lakshit Verma 123196938 MITMPL 30:18
 OK, OK. So just if it works, that's good.
- Adithya Umare 30:33 Ma'am I have a question.
- PO PS 07 30:35 Yeah.
- Adithya Umare 30:38

Ma'am the files that we are looking into are those the real world files which the have the password set by a human or are the password set by computers means completely random like computers have totally random passwords which we can't even remember.

So what kind of passwords are we looking into?

- PO PS 07 31:00
 - So most of the things will be human, human set passwords.
- AU Adithya Umare 31:06
 OK. ma'am OK. Thank you.
- PO PS 07 31:08

 But may be 10% may be completely random also.
- AU Adithya Umare 31:14 OK, ma'am got it.
- PS 07 31:17
 Because sometimes we think it is not, it is random, but sometimes the random files may be having some logic behind it, but it appears to be random.
- Adithya Umare 31:31 Yes, ma'am. Got it.
- PO PS 07 31:38

And to the earlier question, we are raised like you can pre trained model you can have but when you are coming here you have to bring your model and you have to explain this thing how your model is novel and how you have approached what is your approach.

- Raj Mehta 32:01

 Do you have one more hand raise, Dinesh?

 Go on with your question.
- Dinesh Pawara 32:08
 Yes, good morning.
- Po PS 07 32:10 Yeah. Good morning.

- Dinesh Pawara 32:13
 - My question is, what is the scope of this problem statement to crack only just Office 2013 and PDF version 1.7? Why not latest version and above?
- PS 07 32:27
 Yeah, after this, whatever minimum target we are keeping from Office 2013 and above.
- Yeah.

 OK. And yeah, sure.
- Po PS 07 32:36 But I please do.
- **Dinesh Pawara** 32:42 Yeah, you can tell me ma'am.
- PS 07 32:44

 Yeah, means minimum is Office 13 and above. We are not looking into older version.

 We are looking into Office 13 and above and PDF 1.7 and above.
- Dinesh Pawara 32:57

 And what is the link should be it password length?
- Po PS 07 33:02 Password length for stage one is 12 and above.
- **Dinesh Pawara** 33:08
 Okay got it. ma'am. Thank you.
- Po PS 07 33:13 Yes, Ashwin.

- **AN ashwin nair** 33:16 Am I audible?
- Po PS 07 33:17 Yes.
- Raj Mehta 33:18 Yeah.
- ashwin nair 33:18

 OK, so we're using a. Is it important for it to be explainable like why it might get certain password? Or would you guys just care about the performance?
- PS 07 33:31

 No, sorry I could not get you. Can you please repeat?
- Yeah, yeah. So does the model have to be, like explainable? Is that like a measure as to like, how? How did guess that password or it's just performance important like the success rate of guessing, I guess.
- PS 07 33:50

 Yes, success rate is very much important, but at the end, we also question you why you have selected this particular model, why you are not, what is the advantage of this model and how you have approached this model, what is the benefit of it?

 We will be asking you to justify why this model is been selected.
- Sanjay 34:16
 Ashta Ajith here. Can you throw some light regarding the form to which you have submitted? They will have to finally submit that also.
- Po PS 07 34:27 OK, Sir.

Sir, that form it has been published in the site Sir or? It does it to be.

Sanjay 34:35

Not yet published it it it, it will be published by 15th of September, but that is what the final thing is. So I think if we tell means if we clarify now only those are the kind of things which we expect from the.

- **PO PS 07** 34:39
 - Yeah.

Yeah, that is fine.

- Sanjay 34:49 Ah.
- PO PS 07 34:50

OK, Sir. Fine. Fine. I will just let them know that thing. But I think you can publish that earlier also. No need, Sir. OK, Sir, I am just.

- Sanjay 34:59 Yeah, yeah.
- PO PS 07 35:00

Yes, Sir. So when you are submitting your solution, we will be asking you some. We are giving you a format in which you have to submit your solution in that first point will be your executive summary. You have to briefly describe your proposed solution and the key features of your.

Solution. Then your problem understanding like you have to describe how you have understood the specific challenge and the selected problem then some technical overview we will be asking like which a I technique or methodology you are using. And.

Why you choose this model and what is the novality in that? And then we will be asking you your approach for this password recovery. What is the approach you used?

And how your solution has been improved from the traditional methods like brute

force and other things. And further we will be asking you to.

How you handle the complex and long password mix from the symbols and other things and what learning strategy and what air and ML architecture your system you have used.

How does you adopt to the different encryption mechanism when you are changing this file format and we ask about the deployment readiness of the model. Also like your proof of concept prototype market ready.

This thing and security and ethical consideration we will be.

What is the security measure you have taken so that your solution ensures that the integrity and confidentiality of the data being used during recovery so?

And further some other things related to you, whether you have already worked on such type of problem statement and what are the resources required for you.

To do this problem statement like your CPUGPU softwares, data set and scalability and the potentiality and all of your solution. And further we will be asking the details about your team members, your publications and all.

And you have to upload your doctorised model.

Yes, that's it.

Yes, some questions are there. I think you can ask.

Jyotimaya Mohanty 37:40

Yeah, I have a questions now. This will be more specific to user consent or this is a open environment. Anybody can build their module and retrieve the password or you need to have if my platform has created specific.

Consent based approach with a verified identity who can retrieve the password or it open to anybody the model?

PO PS - 07 38:11

Open to no I am not getting you exactly what you want from.

Jyotimaya Mohanty 38:16 So.

Sanjay 38:16

Ashta, I think what he is trying to ask if I may jyothimay you may correct me. Monty

you want to ask this model will be usable by anyone and everyone. So for that the model has to be an offline deployment.

- Jyotimaya Mohanty 38:21 Sure.
- So it is not accessible online at all, so it will be used by the organization only.
- Po PS 07 38:31 Yes.
- Jyotimaya Mohanty 38:31 Room.

OK, OK. Organization or the user who has created the file.

Sanjay 38:41

Basically, basically it is an offline thing. We are not publishing it online so that anyone and everyone will use it. It is not a commercial use of this. When you develop and if you get selected in the final stage and when you actually deploy it, it is going to be an offline deployment.

Of the final model.

- Jyotimaya Mohanty 38:57 OK.
 - Understood. Understood.
- Sw Shubham Wavre 39:04 Take another account.
- RM Raj Mehta 39:09
 We have Aditya again, Aditya.
- Adithya Umare 39:13 Yes, Sir. So my.

- Po PS 07 39:16 Yeah.
- Adithya Umare 39:18
 May I continue?
- Raj Mehta 39:20 Yeah, please.

There, there is some disturbance in your background.

AU Adithya Umare 39:25 Yes.

Yeah, Sir. So my question is, is there a limit to the number of password guesses that there are for each file, or will we be getting the files then we can we download the files, then try on it and then submit the final passwords?

- PS 07 39:45
 Yes, sure. You can download and submit the final password.
- Adithya Umare 39:49

 And while submitting like will that one attempt will be considered the final attempt and will be getting the scores?
- PS 07 39:58

 Yes, like once if you if you have put a right password only your file will open right by that itself. You can be assured that your file you have cracked the password rightly.
- Adithya Umare 40:09
 Right, right. OK, ma'am got it.
- Raj Mehta 40:14

 We have Abhinav with his hand raised Abhinav.
 - Abhinav 40:20

Yeah. So ma'am my my first question is that since level as you said that level one is before before the final submission and level 2 or Level 3 will be after the final submission. So what if like there is there are some changes that in the model that we have to do so that it works smoothly for final submission like.

- RM Raj Mehta 40:20 Go on.
- Abhinav 40:40
 The Level 2 and Level 3 parts. So how will that work?
- PS 07 40:44

 No, no, no, sorry. All level one, level 2, Level 3. After Level 3, only your final submission, we will ask you to submit the final submission first. We will be issuing, giving, giving you some mock test for 50 files to improve your model that we will be giving you giving to you on 15th of September.
- Abhinav 40:54 OK, ma'am.
- PS 07 41:03

 November itself so that you can keep trying and you can time tune your model. Then

we will start releasing our level one files from 26th of October and Level 2 file on first of November and Level 3 files on 6th of November. Everything will be 10:00 AM in the morning so that as soon as you get.

- Abhinav 41:05 OK. Okay.
- PS 07 41:23

 The files you can start cracking the files and you can upload it the upload the time and even mock test files you have cracked. You can give it but mock test files are not evaluated, only level one level 2 level also evaluate.

Abhinav 41:29

What?

Got it.

Yeah, ma'am. Since the final submission is 31st October, if I remember right. So and you are telling that level one and level 2 and Level 3 will be first and 3rd November. So like you are telling that after we have to submit that.

PO PS - 07 41:52

No, no. You can. You can take it till 10th of November and this problem statement Ke Liye till 10th of November. You have you can crack your files and submit, your evaluation will start after that. Also we have some 15 to 20 participants who will be selected. They will be called.

Abhinav 42:08

Got it. Second question, ma'am, since like in few of the other problem statements, they have mentioned that initial initial like initial part of the less stage one and stage 2, they will be providing the open source data or something like.

Like in the models will be mostly open for trained and for like performing on open source data and then the for the stage 3 they will be actually releasing something that real government files and we have to work on those real government files.

And there might be a slight change in those the type of data. So will in this problem statement it will also be the same or the data will remain very similar through stage one to stage 3.

PO PS - 07 42:40

Yes.

Yeah. Here tona. Like when you want to test your model, you can generate your own file and you can protect with your password and you can give it to your model and check whether you are able to break your files right. So here you need not.

- Abhinav 43:08 Yeah.
- PO PS 07 43:08

Squared our files like you can select your own files, you can encrypt it with the password of length 12 and above and you can give it to your model and check it. So here we are not giving you any data set for training and open source data set. You can take it for training your models.

But The thing is, anyhow, we are giving you mocked test files 50 files to before starting our level one test. OK so that you can use it for your trial also like.

- **Abhinav** 43:41
 - OK, so from stage one to stage 3, the file type and the password type and everything should remain fairly constant and we can use our own mob files, right?
- PO PS 07 43:50 Hmm.

So actually, yeah you can this thing, but when you go to stage 3, there is a slight difference in that in stage one you are having passwords of length 12 and above. In Stage 2 you are having passwords of length 14 and above and when you move on to stage 3 it.

Adds on one more feature like multilingual is also getting into it. So it is making your problem little more complicated and in stage 2 apart from Office 2030 we have archives and other things included.

- **Abhinav** 44:19
 - But.

Got it. Got it, ma'am. OK. That's all. Thank you.

- PO PS 07 44:29 OK.
- Raj Mehta 44:31
 Yeah. Next in the Q, Aryan.
- IMT2023548 Aaryan Rajesh Antala 44:34

 Hello ma'am, I wanted to clarify what are the deliverables for the part 1I understood that that there are this mock test and we need to submit a dockerized version. So do

we need to submit the passwords which we cracked for the mock test with the dockerized version of the model?

PO PS - 07 44:51

See. Yeah, you can submit your passwords this thing, but for level one level 2 definitely you have to submit based on that only you will be evaluated as well as you have recovered your password for level one level 2, Level 3 you have to submit the passwords in the portal.

Based on because your time taken also is considered.

So whoever is submitting their time, the time of your submission is we consider as a time taken to crack the password. So as soon as you recover the password at level one, level 2, Level 3, you have to submit it in the port.

- IMT2023548 Aaryan Rajesh Antala 45:26
 OK, so the time at which we submit, it's important, not the time it took to crack the password by our model, right?
- Po PS 07 45:31
 Yes.
 Yeah, yes, yes, sure.
- IMT2023548 Aaryan Rajesh Antala 45:35 OK. Thank you, ma'am.
- Raj Mehta 45:40 Ayush your turn.
- A Aayush 45:43
 Yeah, yeah. Are we expected to handle partially corrupted files or incomplete files?
- PS 07 45:51

 No, we will give you good files, not partial corrupted or some problem files. If you have any issues. Once it is uploaded, you can just immediately tell us we will correct it, but we are we we will ensure from our end.

- A Aayush 46:03 Yeah.
- PS 07 46:07

 The files are not corrupted by chance while uploading some error happened or something you can please let us know but it will not happen 99%.
- Aayush 46:16
 Yeah. So another another question that I have I have is, will we be provided like extra metadata regarding the files to help our models get a bit understanding of the password?
- Po PS 07 46:27
 Wait.

 No metadata. We are not giving at this stage one.
- Raj Mehta 46:42

 OK. We have next Mr Lekhi Reddy, Naveen Kumar.
- LEKIREDDY NAVEEN KUMAR REDDY [CB.SC.P2CYS25014] 46:49

 Ma am my question is for the efficiency metric, will all teams will be tested on the common data set provided by you or an unique challenge files?
- Po PS 07 46:58

 No data set you can take from open source. We are not giving any data sets to work on for this problem state.
- LEKIREDDY NAVEEN KUMAR REDDY [CB.SC.P2CYS25014] 47:04
 OK.
 OK.
- Raj Mehta 47:10 Rahul of metisoft.

Rahul - Matisoft 47:14

Ma'am I want to clarify that we need to take care of the custom encryption also like the extension is DOC X but internal data is custom encrypted.

- Po PS 07 47:24 No, no, we are not giving you any custom encrypted.
- Rahul Matisoft 47:29 OK, ma'am. Thank you.
- Po PS 07 47:39 Ayush.
- Raj Mehta 47:40

 No, Ayush has been done. Ayush, you have some more questions otherwise you can down down your friend here.
- PO PS 07 47:42 OK.
- A Aayush 47:44

 No, no, I don't have any questions.

 Yeah.
- Raj Mehta 47:50

 Is there any other question on this particular new permanent statement 7?

In case you have any question, you can raise a raise an email to at contact us@ai-grand-challenge.in and that shall be replied.

- PO PS 07 48:19 Somebody raise their hand, right them.
- Raj Mehta 48:22

 No, I can't see anything.

- PO PS 07 48:24 OK.
- Raj Mehta 48:27

 So I think this is time to call the webinar off. I thank you all for coming here and putting up your questions. I I also thank Doctor Asal Asalakshmi for giving responses to the applicants. I also thank Kanal Achith for here in the webinar.

 And with this I am going to end the webinar. Thank you very much. Once again. Bye.
- PS 07 48:52 Thank you everybody.
- Meenakshi 48:53 Thank you.
- SN Swaroop M N 48:55 Thank you.
- Lakshit Verma 123196938 MITMPL 48:57 Thank you, ma'am.
- Rutuja_aerovania 48:58 Thank you.
 - Raj Mehta stopped transcription