1. **As the Chief Security Engineer in a new startup that works on autonomous driving, you are responsible for ensuring the safety and security of your company’s ML-based autopilot system, the core product of the company. Answer the following questions.**
   1. What is the threat model for your product? Provide a numbered list and briefly describe each item.
      1. ASSUMPTION: We assume that the autonomous driving deploys a model which is ready for Inference while each user travels inside the car.
      2. DATA COLLECTION STAGE- DATASET(BOTH DIGITAL AND PHYSICAL):
         1. As the user travels inside the car, the sensors can pick up information by mistake or a bias could be introduced intentionally into the sensor  
            Forcing the car sensors to think that there are no other cars nearby or about a faulty condition of roads/ direction during commute, the adversary can trick the model to make erroneous decisions
      3. TRAINING ATTACK:
         1. DATA POISONING: adversary introduces samples which could perturb the decision boundary during the training, leading to poor, insecure performance during the online.
         2. BACKDOOR ATTACK: If we assume that the user somehow gains access to the model during training stage, then he can launch a backdoor model using particular datasets to embed  
            backdoor features into the model as the model gets trained, this will make the model make intentional inaccurate predictions while going into the online mode.
         3. GAN ATTACK: Malicious entity can use GAN attack during training itself and try to reproduce the training network as closely as possible without even knowing too much details about the model
      4. INFERENCE ATTACK: Hence we can think that Model Inference is happening inside each user's car and each user and any entity associated with car has access to the model.
         1. EVASION MODEL- MODIFYING INPUTS DURING INFERENCE-: Add minor modifications to the input samples during inference such that the model classifies the Modified, perturbed input sample wrongly.
         2. STEALING SECURE BLACK BOX MODELS: User can be Honest/ Non Malicious or Malicious,   
            If the User is Malicious, the user can use the model deployed in the car and perform Model Stealing/Extraction Attack, User can query the original model with his dataset,   
            Use the Original Model's prediction outcomes to train another shadow model   
            User can run an algorithm which can continuously compare the predictions of both original and shadow model, use a metric to see how closely Shadow/Malicious model tracks the Original Model
         3. GRAY BOX MODEL: User Access Privilege to Model- Depending upon the user's accessibility/model encryption, the user may try to style Model hyperparameters or weights by crafting attacks against such secure layers
      5. PRODUCTION/INTEGRATION:
         1. AI Bias- The model could be trained to inherently possess bias against specific Roadmaps, Specific kind of traffic, specific crowd comprising of different race/sex.  
            Models in cars could be tricked to differentiate between these different entities and take specific actions in each case, such as slowly down a car in a Roadmap comprising a shady area, accelerating car in a location busted with traffic or behave differently when the car witnesses people from specific genders.
         2. We know that many ML services also use some limited open source API's and resources to deploy their specific applications, a malicious user can target such open source API's and introduce modifications implicitly through the loss or optimizer (which could pass Pull Request)   
            Such modifications could introduce major modifications during training of the model while the computational graph gets build up, which could contain specific malicious components forcing inaccurate/misleading outcomes at the end
   2. For each item above, discuss what are the potential solutions.
      1. SOLUTIONS
      2. DATA COLLECTION STAGE-
         1. STATISTICAL ESTIMATION, DETECTION, TRAINING:  
            Users can use advanced methods to generate statistics which capture the important characteristics/distributions of the dataset, while the model is used online,   
            Users can use this as a reference to do data validation before making inferences on datasets blindly
         2. DATA MANAGEMENT:  
            Have well defined clean practices of managing data which would be trusted and relied upon during training by secure practices, the database could serve as permanent reference for future revisions and to compare against emerging datasets.

1. TRAINING STAGE-:
   1. ROBUSTMODELS/TRAINING:   
      Use training techniques which converge to models which have robust, constrained decision boundaries to ensure that the model is not susceptible to data poisoning attacks
   2. CERTIFIED DEFENSE:Use certified defense techniques which are proven safe through years of practice and deployment
2. INFERENCE STAGE:
   1. TRAINING MANY ENSEMBLE MODELS:
      1. Train multiple models parallelty, so that you can develop confidence levels of your prediction accuracy against all the other models and Not relying on the output of a single model leading to erroneous decisions/labelling/misclassifications
   2. VALIDATION NETWORK:
      1. Have a validation network which verifiess your prediction after inference before the systems release predictions,
      2. If the predictions are disparate, there can be algorithms to figure out which model is more faithful in classifying the input sample
   3. MODIFIED NETWORK
3. PRODUCTION/INTEGRATION
   1. MODEL WATERMARKING- is a potential solution to always ensure that what gets shipped is totally verified and foolproof
   2. SECURE CODE REVIEWS AND UNIT TEST BASED INTEGRATION:   
      Developing unit tests which are robust against production layer level attacks could be incorporated In the open/closed source code base repositories
   3. DIFFERENTIAL PRIVACY- Technique could also be used

1. For each solution described above, discuss the overhead (e.g., design cost, performance overhead, energy overhead, time-to-market delay, etc.).
   1. DATA COLLECTION STAGE- STATISTICAL COMPUTATIONS-   
      Design need not be altered, but there could be additional cost incurred, energy overhead involved in performing the computations to arrive at statistics which is still robust to errors and can be sent to market with trust passing all certifications
   2. TRAINING STAGE-
      1. ROBUST TRAINING:  
         Again there is additional compute energy, lateny cost in training more networks to perform such detection, This once done, could be reduced and be sent to production ready stage products.
      2. CERTIFIED DEFENSE:  
         Time to market delay is the major deterring factor for obtaining certified defense, as we would have exchange Ips ,get our designs evaluated by governing authorities, iteratively carry out their revisions if any to successfully get certifications to ship the products.
   3. INFERENCE STAGE:
      1. ENSEMBLE MODEL- generation is memory intensive process and may not be viable for resource constrained edge devices, compute and latency costs are incurred one time but the energy overhead stays throughout as you have to validate the decisions generated by running all these models
      2. VALIDATION NETWORK: This could be lightweight in all aspects compared to the first solution, but its robustness is also lower compared to the first one, as you have just a single validation network which could also be wrong at times and propagate its biases to the original model
   4. PRODUCTION/INTEGRATION:
      1. MODEL WATERMARKING: Model watermarking has design overheads too and hence might be difficult
      2. SECURE REVIEWS: This delays the time to market time without incurring any other resources.

1. (After a few months…) Because of your excellent vision and hard work, you have been promoted to CEO (Congrats!). For an early-stage startup company like yours, discuss how you would decide which security solution to use and what to discard. In other words, what would be your strategy to provide a balance between security and overhead?
   1. I would opt for solutions which provides an optimum tradeoff of security and overhead without delaying the time to market which is valuable.
   2. I would try to advise my company to invest a lot of time in having a highly robust Trusted Computing Base and do not touch it back again in the near future.
   3. Having a reliable TCB, I would encourage multiple teams to continue to build multiple layers or tiers of security in a decoupled manner, build fault tolerant systems for backup.
   4. I will try to stick to Proven designs and would not pick choices which need the design altered as it may lead to deterioration of product quality if the added modifications do not work well and diminish the product quality. This also improves the speed to reach out to markets.
   5. I will operate based on continuous feedback from users
   6. With the obtained feedback, I would add several iterative versions of lightweight, proven secure modifications which could be adopted within the compute resources that the users might already have.

1. **In the class, we discussed an informal solution for Yao’s Millionaire problem. Describe how this problem can be solved using Garbled Circuits (GC). Assume that Alice and Bob’s inputs are {1 m, 2 m, 3 m, 4 m}**
   1. First, provide pseudocode for Oblivious transfer, and call that OT(m0, m1).
      1. Alice has two messages m0, m1 and Bob wants to know one of them.
      2. Alice creates two random value k, on the basis of the message, he wants and sends v=enc(k)+xb
      3. Alice computes k1=dec(v-x1) and k2=dec(v-x2), then sends t1=m1+k1 and t2=m2+k2
      4. Bob knows which ki is valid and computes mb=tb-kb
   2. Provide step-by-step pseudocode for GC for a 2-input NAND gate, and call that GC(a, b).
   3. Chart

      Description automatically generated
      1. Alice and Bob wants to find the output
      2. Alice has input g, Bob has input e
      3. First step is garbing
         1. Create a random label for each inputs Wi that Alice and Bob have
         2. Encrypt the output based on the labels which is a two-key encryption
         3. Text, letter

            Description automatically generated
         4. The garbler then creates a random input permutation
         5. Text

            Description automatically generated
      4. NAND Table:
      5. Let's assume Alice does the garbling, she then sends the table to Bob
      6. She also sends her the random lookinng input
      7. Bob now needs to do the evalutaion
      8. Bob has his random label We but also wants Alice's random label Wg,
      9. Bob achieves this through a secure protocol called oblivious transfer
      10. Bob now has both keys /labels, evaluates all the rows and sends the values,
   4. **Provide step-by-step pseudocode for solving the Millionaire problem using (a) and (b**
      1. NAND gate can be used to build any logic as it is a universal gate
      2. Build Compartor circuit using NAND gates available
      3. Truth table for comparator is evaluated, it is then garbled by Alice as mentioned above and sent to Bob
      4. Comparator will be used to compare the wealths of both Alice and Bob
      5. Alice also sends the key that she has
      6. So Bob needs the other key to compute answers to the table
      7. For the exchange of the other key, we have oblivious transfer algorithm to let two parties securely exchange data without mutually knowing each other's original input or data.
2. **After successfully selling your autonomous driving company to Ford (Congrats, you are a millionaire now! How much did you earn, BTW? Is it more than me?), you decided to start a new company to compete with Amazon Go. Your main product is the technology for enabling a cashier-less grocery store**
   1. Due to privacy concerns, training your ML system is a major challenge. Customers do not want to be identified and/or tracked (their face, their shopping habits, the items that they purchased, etc.). Describe how would you address this issue?
      1. There are two potential solutions to this problem  
         **ARL- Adversarial Representation Learning and HE - Homomorphic Encryption  
         CryptoNets can also be utilized and even improved models such as LoLa can be utilized**Overhead- HE- Compute resources, power and most importantly latency,   
         To reduce overhead of homorphic encryption, we need a complete set of operations over packed vectors of plain texts  
         Using Leveled Homomorphic Encryption technique to know the arithemtic circuit complexity before entering into the computation part.  
         Also we can use CRT Theorem to encode large numbers, exploit parallelism with SIMD instructions to improve latency  
         HE- can't handle non-linear functions, need special ways of approximating non-linear functions with lower order for polynomials
      2. I will chose LoLa with robust Homorphic encryption- along with some online offline training techniques to further reduce the workload
   2. What is the threat model?
      1. Membership Inference Attack:   
         Adversary can launch membership inference attack that is given access to a Black Box Model's prediction API, attacker could determine if the record was in the model's training
      2. Use of Backdoors to exploit sensitive attributes:  
         Honest But Curious Classifiers can always infer additional sensitive attributes and it is reported that it is impossible to censor sensitive atttributes even though they are intentionally introduced as part of the training data or not part of the training data itself
      3. Model Inversion: Access to a model is abused to infer information about the training data

1. **If you are into Wordle these days, try to solve this puzzle:** [**https://enigmadevice.com**](https://enigmadevice.com) **(easy version).** 
   1. Why, sometimes I h've believed as many as six impossible things before breakfast   
      --Alice in Wonderland   
      Date- 02/05/2022
   2. Answer-   
      The disadvantages found include the main flaw of the machine itself, that the machine never encrypted a letter to itself, meaning that the use of cribs was possible and furthermore group theory.
   3. As well as this, the report exposes the operator errors that led to Enigma being broken.  
      Approximation to alphabet/ rotor start positions

Starting with a rotor order from step 1 and w, trying all rotor positions and ring positions for 1st ring only, again searching highest IC. That takes 264≈219 operations.

Find ring and rotor start positions

We have the 1st ring and 1st rotor start positions, approximations for other rings and starting positions from 1 and 2. First search positions for the 2nd ring and rotor, then use the same procedure for last remaining rotor. This will take 262≈29 operations.

Find the plug settings

We have the rotor order, position and ring positions. Now we can use IC as statistical test again, deriving of the trigram information of the underlying language.

1. There are a variety of semanticallty secure algorithms which we can apply here Goldwasser-Micali, Elgamal , Paillier.   
   Other semantically insecure algorithms such as RSA can be made semantically secure through the use of random encryption padding schemes such as Optional Asymmetric Encryption Padding (OAEP)
2. **5. So far, what did you like about the course, and what did you don’t like? (lectures, material, format, workload, etc.)**
   1. I like all the lectures and materials in this course,
   2. I extremely liked the emphasis given on fundamental concepts rather than the deeper math behind them.  
      Especially, the way Professor easens the crypto load and explains all the concepts in a very intuitive way with various real life exmples is awakening and exiciting,
   3. While also, not overwhelming us with alll the complicate crypto abstractions
   4. Workload - reviewing 4 papers per week is little difficult, but to my surprise, I found that I learn a lot more  
      If I go through the papers before class presentations
   5. Also, I like the way how correlated papers are organized and connected together
   6. So we have a well defined mind map of concepts required to tackle the paper reviews and necessary math rigour to understand lectures while in class
   7. The project is also insightful as we get to survey across a variety of topics