

# Getting a job in AI in 2025

A practical course

A dark blue diagonal gradient bar that starts from the bottom left and extends towards the top right, covering the lower half of the slide.

# Why should you consider Data/ML as a career path?

- **More future-proof**



- Higher barrier of entry.
- ML requires deeper knowledge than SWE.
- Unless self-improving AI happens in the near future, there will still be a need for Data/ML experts to advance this technology.

- **Higher compensation**



- When I graduated UBB (2020) and then Edinburgh University (MSc AI), distributed systems eng and ML eng were the highest paid in tech.

- **There is a lot of capital poured into AI**



- Everyone is investing in it (including governments), YCombinator recently upped their funding batches from 2 to 4.
- Even if all funding stops tomorrow, there are a lot of jobs at startups that already got the money for a solid 10 yrs+ runaway

- **Working at the most innovative/coolest tech**

# Why should you trust what I say?

- I am an ML engineer based in **London** with **5+** years of experience and 7 academic publications with nearly 100 citations total.
- I am a former hiring manager that conducted over **100** ML and Data Science interviews across Europe and North America.
- The roles I was hiring for offered base salaries ranging from **\$13k-\$20k/month**. This background has given me a clear understanding of the calibre of talent expected at the high end of the market as well as insight into the skills employers truly value.
- I also designed a substantial collection of original, technical interview questions that I routinely used to assess candidates during these sessions.



# Understand the market and the requirements.

You have to show proof of exceptional ability.

## Data/ML

- Usually requires at least Msc
- Cool and complex personal projects (AI related)
- Good thesis/dissertation
- Academic publications (be it only Arxiv, still counts). Bachelors is ideal to get your 1st publication
- Internships in AI
- Capable in math/other sciences is a big plus.
- ML / Data roles tend to be more elitist (unfortunately)

## SWE

- BSc is usually enough
- Cool personal projects
- Internships in SWE
- Strong competitive coding / DSA skills (leetcode).

# What is achievable as a new grad in Data/ML?

## New grad

Requires:

- Usually MSc but sometimes doable with BSc only
- Cool and complex personal projects (AI related)
- Good thesis/dissertation

Expected compensation:

~\$50k-80k/yr

## Junior

Requires:

- Usually MSc but sometimes doable with BSc only
- Cool and complex personal projects (AI related)
- Good Internships in AI

Expected compensation:

~\$50k-120k/yr

## Mid-level

Requires:

- Usually MSc
- Cool and complex personal projects
- Some academic publications
- Capable in math or other sciences

Expected compensation:

~\$120k-250k/yr

## Unicorn

Requires:

- Phd from Top 10 Uni or IMO/IOI/IMC (competitive background)
- Publications or extremely good personal projects.

Expected compensation:

~\$150k-400k/yr



Achievable straight out of Uni, but only in Hedge Funds (London/New York/Chicago)

# I want to apply to a job tomorrow, what do I do?

1. Brush up your Resume / CV
2. Make a portfolio of 1-3 complex personal projects in ML (if you don't have already)
3. Look to publish your thesis / dissertation as an actual academic paper (Arxiv is good enough for now).
4. Prepare for interviews
  - a. Statistics
  - b. Classical ML theory
  - c. Mathematics (calculus / linear algebra)
  - d. Some leetcode for data manipulation (e.g. string manipulation exercises) are useful too



# 1. Resume. What I used to get interviews at the highest level e.g. Jane Street.

Resume should have these core sections:

1. Name and contact info
2. Education
3. Work Experience (if any)
4. Awards/Accomplishments (if any)
5. Technical skills (programming languages / technologies)
6. Personal projects (very important)
7. Publications (if any)
8. Extracurricular Work
9. Other skills (e.g. hobbies, english proficiency)

green - core

blue - good to have

Header: Name  
and Contact Info

Work experience

(try not to leave empty,  
put any internships)

Pricope Tidor-Vlad

✉ ptidor1@gmail.com | ☎ +447909820976 | 🌐 TidorP | 💻 ptidor.com | 📝 blog

## Education

- **The University of Edinburgh**, Scotland, UK. | 2020 - 2021  
MSc in Artificial Intelligence. Coursework avg. **83.23/100**. GPA: *First class* (4.0).
- **Babes-Bolyai University**, Cluj-Napoca, Romania. | 2017 - 2020  
Bachelor's degree in Computer Science. GPA: **9.87/10**. Thesis: **10/10**.

## Work Experience

- **Machine Learning Engineer Lead - Ntropy** | August 2021 – April 2024
  - Fintech company series A, developed and deployed language models for financial transactions understanding.
  - Some of my work here is public and can be viewed on [this](#) medium post (co-author).
  - Hiring operations – interviewed over 100 candidates for machine learning and data science roles at Ntropy.
  - Fine-tuning and optimizing LLM usage through smart caching was also a big part of my work.
  - Tools used: *transformers in pytorch, onnx+triton, fasttext, sklearn, multi-gpu training, git, jupyterlab, aws, LLM*.
- **Quantitative Research Assistant - Predictiva** | April 2021 – August 2021
  - Developed Deep Reinforcement Learning models for portfolio management for algorithmic trading.
  - Tools used: *python, RLlib, tensorflow keras, cloud computing*.
- **Quantitative Research Intern - Predictiva** | February 2021 – April 2021
  - Developed Deep Reinforcement Learning models for quantitative stock trading.
  - Tools used: *python, cloud computing*.
- **Udemy Instructor** | June 2020 – August 2020
  - I devised and published 2 AI courses. I had about 40 students on Udemy and 15 on Skillshare, with a rating of 4/5.
- **Autonomous Driving Research Intern - GmbH Robert Bosch** | July 2019 – Oct 2019
  - Developed optimization and simulation frameworks for camera windscreen distortions. Computer Vision.
  - Tools used: *matlab, projective geometry, multivariable calculus, optimization through adaptive algorithms and ML*.

## Awards

- Second place at the **IBM MLP Competition** out of 101 projects. Project subject: AI Quantitative stock trading. 2021.
- Two Silver Medals at **South Eastern European Mathematical Olympiad** for University Students 2018-2019.
- Two Silver Medals in **top 10** and a bronze medal in **top 20** at **National Mathematical Olympiad** 2014-2016.
- Member of the **extended National Senior Team** for the **International Mathematical Olympiad** 2014.
- Diploma from **Babes-Bolyai University** for academic performance 2018.
- Top 20 at **National Chess Olympiad**, 2017.
- Other 70+ prizes in math and chess international and national competitions 2005-2019.

Education

(more important when  
early in career)

Awards

(anything counts)

# 1. Resume. What I used to get interviews at the highest level e.g. Jane Street.

Resume should have these core sections:

1. Name and contact info
2. Education
3. Work Experience (if any)
4. Awards/Accomplishments (if any)
5. Technical skills (here you write programming languages / technologies)
6. Personal projects (very important)
7. Publications (if any)
8. Extracurricular Work
9. Other skills (e.g. hobbies, english proficiency)

green - core

blue - good to have

## Technical skills:

languages/technologies

## Publications

## Other skills

### Technical Skills

- Data Science: **Pytorch**, **Tensorflow**, **Keras**, **Onnx**, **Triton**, **Scikit-learn**, **Pandas**, **Polars**, **Numpy**.
- Programming Languages:
  - Very confident with: **Python**, **Bash/Shell Scripting**.
  - Comfortable with: **C++**, **SQL**, **Matlab**, **Java**.

### Projects

#### Personal Projects

For more projects or additional information check my [GitHub](#) page or my [website](#)'s *projects* and *blog* sections.

- **Poker AI** – I developed 3 Poker agents, the best one using *Deep Reinforcement Learning* and achieving expert human level. I published [a paper](#) related to this project and [a paper](#) is waiting for publication - [poker.pttdor.com](#)
- **Quant-Trading AI** – a project that uses deep learning and technical analysis on stock trading. I developed expert-system agents trained to trade on 20 different stocks; 17 were successful with best return of 400% over 6 years.
- **Portfolio Management with DRL** – my dissertation project with the scope of building a Deep Reinforcement Learning agent with a new reward function to optimize the profits and diversification on a portfolio of stocks.
- **Virtual Assistant for Students** – a *NLP* project similar to Siri/Google Assistant with extra special features targeted for an academic environment (for example - a course recommendation system based on text or audio input).

#### Publications – 71 academic citations – all single author

[Google Scholar](#)

- **An analysis on very deep convolutional neural networks: problems and solutions.** Published in the international journal *Studia UBB Informatica*, Vol 66, No 1, July 2021, indexed in *Mathematical Reviews*. Single author. [Link](#).
- **Deep reinforcement learning in quantitative algorithmic trading: a review on Arxiv.** 2021. Single author.
- **A view on deep reinforcement learning in imperfect information games.** Published in the international journal *Studia UBB Informatica*, Vol 65, No 2, October 2020, indexed in *Mathematical Reviews*. Single author. [Link](#).
- **A math issue** published in the *Crux Mathematicorum* international journal; volume 44, no. 8, problem 4377 October 2018. Co-author. [Link](#).
- **Subject 3** at the first Romanian Team Selection Test for **Junior Balkan Mathematical Olympiad** 2015. [Link](#).
- **Deep Reinforcement Learning from Self Play in no-limit Texas Hold'em Poker.** Published in the international journal *Studia UBB Informatica*, Vol 66, No 2, 2021, indexed in *Mathematical Reviews*. Single author. [Link](#).

### Extracurricular Work

- In May 2021, I was invited to hold two 2h lectures on Reinforcement Learning for **Babes-Bolyai University**.
- Over the years, I devised and proposed over 40 math problems for the **Romanian Mathematical Society**.
- In my free time, I work on a personal [blog](#) on diverse AI topics and interview preparation materials.
- Started a local Chess Club back in 2013 to help teenagers get better at chess.

### Other Skills

English professional level: IELTS Certificate 7/9 (C1). **Hobbies:** astronomy, competitive chess, table tennis, gaming.

## Projects

## Extracurricular



## 2. Portfolio of cool personal projects

When you are just starting out in your career, employers don't have much to look at in your profile. You have to show proof of practical experience somehow, and projects are perfect to do that.

- focus on projects with a wow effect, e.g. don't pick a generic "sudoku solver" project or a generic CRUD app, solve a real-world problem like image segmentation or real-time classification with ViT or EfficientNet
- complexity needs to be decently high, use cool novel ideas that you can talk about and impress
- use technologies that are or could be used in production systems, e.g. use **polars** instead of **pandas** to handle large tabular data, use modern monitoring systems for training (e.g. tensorboard/WanB instead of manual plots)
- the idea is to show that you can handle complex projects and you have worked with production technologies so that the employer can trust you that you can do good work
- **Always showcase your work on github (very important). You can't apply to a job without a GitHub!**

### 3. Try to get publications (easier than you think)

In computer science, people tend to have a dissertation like an “app” that doesn’t really answer a research question.  
**Be very careful!**

However, if you have been guided by the right people, it is likely that you have a proper potential academic paper that you can write, from your thesis.

Follow these steps:

1. Discuss with your professor the possibility of converting the thesis to a paper
2. Cut the unnecessary details
3. Make a 5-15 page draft preserving the problem statement and important findings
4. Take advantage of the Uni email address and connections to professors to receive the right to publish on Arxiv (you will have it forever afterwards).
5. Optionally, if the paper is super strong, consider publishing it to an actual journal (after ArXiv).

## 4. Prepare for interviews

Since you will likely be applying to new grad / junior positions, you won't receive hard **ML system design** or **big data** interview questions. The interview questions at this stage are more **academic** or aimed towards testing **raw intelligence** (since they cannot test actual **work experience**).

You will likely receive questions about topics studied in school in:

- a. **Statistics**
- b. **Classical ML theory** (decision tree, linear regression, neural networks, data preparation)
- c. **Mathematics** (calculus / linear algebra)
- d. **Leetcode / project Euler** - data manipulation problems. *Some companies still resolve to leetcode, even for ML.*
- e. **Brain teasers**

**Important!** There are a few questions that are asked 95% of the time during an interview (even for seniors). Be sure to have the answers to these questions ready beforehand:

- tell me about a project at your previous job (or internship).
- tell me, in detail, how you developed this personal project.
- tell me in detail about your dissertation, what methods did you use to solve the problem

Be sure to revise these materials ^ before going to an interview.

## 4. Prepare for interviews. Sample questions.



Bosch Gmbh, Cluj-Napoca

\$30k-60k/yr

1. [Bosch] Say you have a classification problem and the associated training set file is sorted by class. You use stochastic gradient descent and you observe that the results are not good. What are possible causes of this, and what would you do to solve the problem?
2. [Bosch] Why do we need gradient descent instead of just taking the minimum of the N dimensional surface that is the loss function?

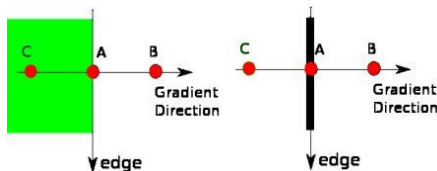
## 4. Prepare for interviews. Sample questions.



BOSCH

Bosch Gmbh, Cluj-Napoca

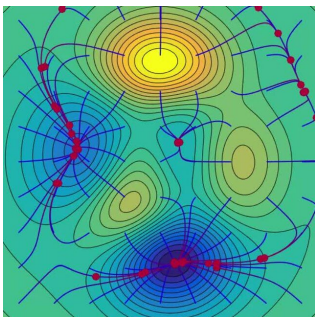
\$30k-60k/yr



one class  
(at a time)

**A short satisfactory answer to this question:**

- This is the hypothetical worst case scenario for SGD.
- The gradient will move in the direction of one class and then the other and most likely miss the global minima
- Solution is to shuffle the data points to increase the chances of SGD converging to the global minimum.
- Another solution is to switch to Gradient Descent, as the sorting doesn't matter there (we compute the true gradient on everything). This is assuming the data is not too large and we can feed in the whole training set.



optimization  
for each class  
at a time

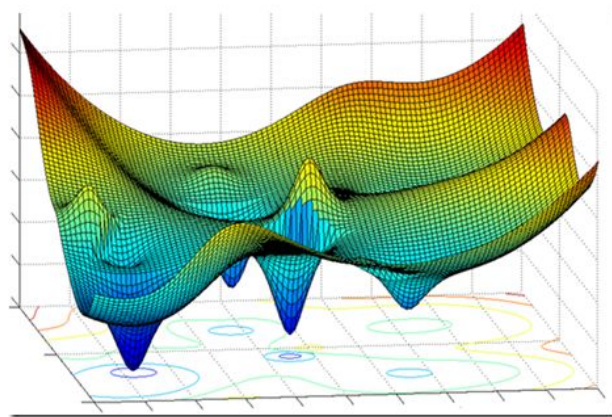
# 4. Prepare for interviews. Sample questions.



Bosch Gmbh, Cluj-Napoca

\$30k-60k/yr

[Bosch] Why do we need gradient descent instead of just taking the minimum of the N dimensional surface that is the loss function?



loss landscape

easy to visualize in  
2D or 3D

a 3D loss surface  
implies a 2-parameter  
neural network!

impossible to visualize in  
high N dimensional

## Ideal answer:

- For some limited problems (like the one on the left), it is indeed possible to analytically solve the differential equation to find the critical points of minima.
- However, real-life problems are too complex and an analytical solution doesn't actually exist, hence we employ an iterative approximate solution through gradient descent.
- Most real problems are also highly non-convex.

# 4. Prepare for interviews. Sample questions.



Microsoft

\$30k-100k/yr

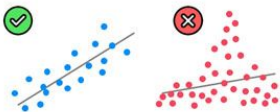
1. [Microsoft] Think about the assumptions of linear regression. If each assumption is violated, what are the remedies?

## Assumptions of Linear Regression



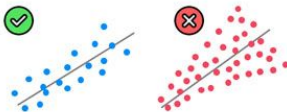
### 1. Linearity

(Linear relationship between Y and each X)



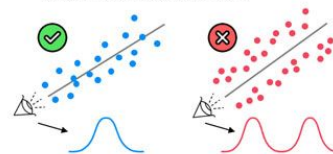
### 2. Homoscedasticity

(Equal variance)



### 3. Multivariate Normality

(Normality of error distribution)



### 4. Independence

(of observations. Includes "no autocorrelation")



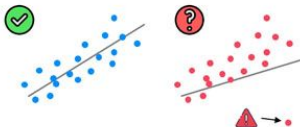
### 5. Lack of Multicollinearity

(Predictors are not correlated with each other)



### 6. The Outlier Check

(This is not an assumption, but an "extra")



# 4. Prepare for interviews. Sample questions.

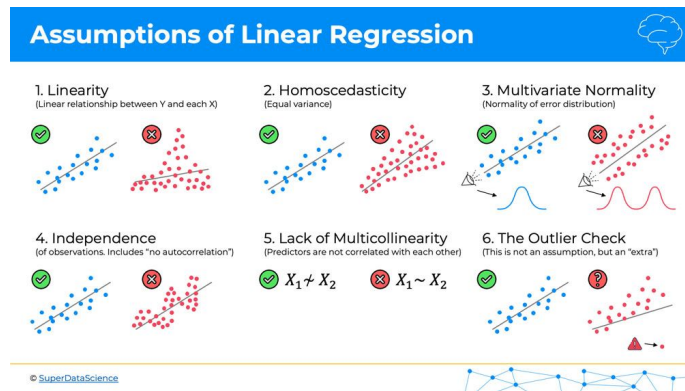


Microsoft

\$30k-100k/yr

## 1. [Microsoft] Think about the assumptions of linear regression. If each assumption is violated, what are their remedies?

- **Linearity:** this can be sometimes achieved (but not all the time) through transformations: (e.g., log, square root, or polynomial) to either the predictors or the dependent variable
- **Independence:** Generalized Least Squares (GLS)
- **Homoscedasticity:** Weighted Least Squares or use heteroscedasticity-consistent standard errors (e.g., White's standard errors) to adjust inference.
- **Normality of Residuals:** transformations (transform the dependent variable (or even predictors) using methods like Box-Cox or log transformation) or bootstrap methods (use resampling techniques to obtain robust estimates of standard errors and confidence intervals)
- **Lack of Multicollinearity:** Remove or Combine Variables (eliminate redundant predictors or combine correlated variables into a composite index) - this can be done by PCA, Regularization Techniques (use methods such as ridge regression or LASSO that can mitigate the effects of multicollinearity by penalizing large coefficients)





## 4. Prepare for interviews. Sample questions.



Microsoft

\$30k-100k/yr

1. [Microsoft] What could happen if you try to fit logistic regression to a perfectly linearly separable binary classification dataset. What would you do if given this situation, assuming you must preserve logistic regression as the model?



Short answer:

- well known problem, coefficients risk going to infinity
- add regularization to tame down the coefficients

# 4. Prepare for interviews. Questions left to reader.



Altametrics, Bucuresti

\$50k-60k/yr

- 1: What is cross-attention in transformer models, and what is its purpose?
- 2: What is overfitting, and how can it be prevented?
- 3: What is cross-validation, and how does it improve model evaluation?
- 4: What is gradient descent, and why is it important in training machine learning models?
- 5: What is the bias-variance tradeoff?

BONUS Question: In the context of OCR (Optical Character Recognition),  
how does a sequence-to-sequence model handle variable-length text in images?

## Question 2 out of 2

Please answer the following questions, including as much detail as possible. Our team will place importance on answers that are concise and easy to understand. Additionally, please answer these questions using examples from your own experience, discussing your personal work related to each topic mentioned. This will give the Altametrics hiring team a clearer picture of your personal exposure to machine learning.

```
1 # Question 1: What is cross-attention in transformer models, and what is its purpose?
2
3 # Question 2: What is overfitting, and how can it be prevented?
4
5 # Question 3: What is cross-validation, and how does it improve model evaluation?
6
7 # Question 4: What is gradient descent, and why is it important in training machine learning
8
9 # Question 5: What is the bias-variance tradeoff?
10
11 # BONUS Question: In the context of OCR (Optical Character Recognition), how does a sequence-
12
```

Submit to see if your code passed

Submit

Skip

## 4. Prepare for interviews. Questions left to reader.



Ntropy (my previous company), Remote


\$150k-240k/yr

### ML live coding interview Sample

1. Take the number 1243. Sort this number in ascending and descending order with respect to the digits. Subtract the lower number from the highest. Repeat the process. My claim is that we are going to converge to 6174 in the end for every 4 digit number (without all digits the same). Make a program to confirm this, it should output the number of steps necessary to reach 6174 (Kaprekar constant). If successful, I claim that for all the 4 digits numbers, we are going to reach 6174 in at max 7 steps, adjust the program to prove that. If successful, I now want a hashmap having as key the number of steps and as value the number of numbers for which we have key-number steps to reach 6174. **Should be max 20 lines of code.**

2. **Sklearn exercise.** Take any binary classification problem (download everything from sklearn). Plot the ROC curve. I am interested in seeing what transformations on the scores (probabilities) won't change the ROC curve. Let the candidate answer. Suggest sqrt to apply -> ROC shouldn't change, make him implement the change. Then ask for a function that doesn't change (sin) and ask him to implement. **Should be max 20 lines of code.**

# Interview questions at the highest level

 **r/MachineLearning** • acum 15 zile  
Arqqady

**[D] POV: You get this question in your interview. What do you do?**

Discussion

(I devised this question from some public materials that Google engineers put out there, give it a shot)

**Q.** [Google Deepmind] An AI company just shipped a new foundational language model. They claim they have trained it for 2.79M H800 hours on 14.8T tokens. Upon further research, looking at Nvidia card specs, you find 3,026 TFLOPs/s of FP8 performance with sparsity, or typically half this ( $1.513 \times 10^{15}$  FLOPs/s) without sparsity. Moreover, you find out that they used FP8 FLOPs without structured sparsity. Given that the model has 37B activated parameters, roughly what hardware utilization did they achieve? Select the closest.

- ☐ 21.7%
- ☐ 16%
- ☐ 28%
- ☐ 88.5%

*Note: For this exercise, you can assume the architecture of a transformer where each of the 37B parameters incurs 2 FLOPs in the forward pass, 2 FLOPs in the backward pass and 2 FLOPs in the optimizer update.*

541

110

Distribuie

362K de vizualizări

Explorează mai multe statistici

Adaugă-ți răspunsul

**r/MachineLearning** S-a alăturat

**Machine Learning**  
Beginners -> /r/mlquestions or  
/r/learnmachinelearning , AGI ->  
/r/singularity, career advices ->...

Arată mai mult

Creată 29 iul. 2009

Public

3 mil.

Members


125

Online

Primii 1%

Clasifică după dimensiune

REALIZĂRILE COMUNITĂȚII

 Picasso, Înțelept

2 deblocate

Vizualizează toate

REGULI PENTRU

1 No Spam

2 No Self-Promotion

3 No Marketing Campaigns (SEO)

4 No Disrespectful Behavior

5 No arXiv Links without Body

## 4. Prepare for interviews.

More questions left to reader.

1. **[Google]** Describe a scenario where base gradient descent is likely to fail in finding a global minimum. How might you overcome this limitation?
2. **[Teambld]** Using a regularizer in a regression setup let's say L1 or L2, how does it affect the minimizer? Do you get the true minimizer? If the base problem is not ill posed, is it possible to get a loss value for ridge regression lower than that of the linear regression loss function?
3. **[Folclore]** Alice recommends the use of convolutional neural networks instead of fully-connected networks for image recognition tasks since convolutions can capture the spatial relationship between nearby image pixels. Bob points out that fully-connected layers can capture spatial information since each neuron is connected to all of the neurons in the previous layer. Both are correct, but describe two reasons we should prefer Alice's approach to Bob's.
4. **[Startup]** You design a fully connected neural network architecture where all activations are sigmoids. You initialize the weights with large positive numbers. Is this a good idea? Explain your answer.
5. **[Royal Bank of Canada]** How would you determine that your dataset follows a Gaussian distribution? What if the data is univariate vs multivariate?
6. **[Uber]** How would you simulate the roll of a fair six-sided die using  $U(0,1)$  (uniform distribution) random number generator? How would you validate that the rolls are indeed fair?
7. **[Apple]** Explain the concept of feature engineering and its importance in the context of tabular data with XGBoost as a predictive modeling tool.
8. **[OpenAI]** How would you design a Human-In-The-Loop (HITL) system to collect explicit labels for a supervised learning task, balancing the need for high-quality annotations with the constraints of cost and scalability? Discuss the methods you would use, the advantages and disadvantages of HITL, and how you might leverage large language models to enhance the labeling process.
9. **[Affirm]** You use a random forest for a classification problem because it is generally a good learning technique which provides good results. However, it seems the model is overfitting on the training data. What do you do?

# Closing thoughts.

## Where should you apply?

**Top companies in Romania for total comp**

(>\$9k/month at mid-senior level)



Bucharest: CrowdStrike, Bolt, Adobe, Emag, top betting companies.

Cluj (\$4k-\$9k/mo): Bosch, Yardi, Accesa, Steelcase, Betfair.

Or...

**Contractor for a US-based company**

(>\$10-20k/month at mid-senior level)

American startups (ideally YC backed), medium sized companies that hire remote.

# If applying abroad



NETFLIX



Uber



The above are the best paying companies in ML / Data space. Their total compensation for mid-senior level goes to **\$25k-50k/mo** on average, with the top 0.1% engineers making **>\$1M/yr**.

If you are ambitious and you work in tech, you should move to

**San Francisco**



It's the undisputed #1 place in the world for growth and opportunity in tech.



# [neuraprep.com](https://neuraprep.com) – an interview prep platform for ML/DS

Most SWE go to [leetcode.com](https://leetcode.com) and learn DSA to land a job. However, there isn't an equivalent place for preparing for Data/ML interviews!

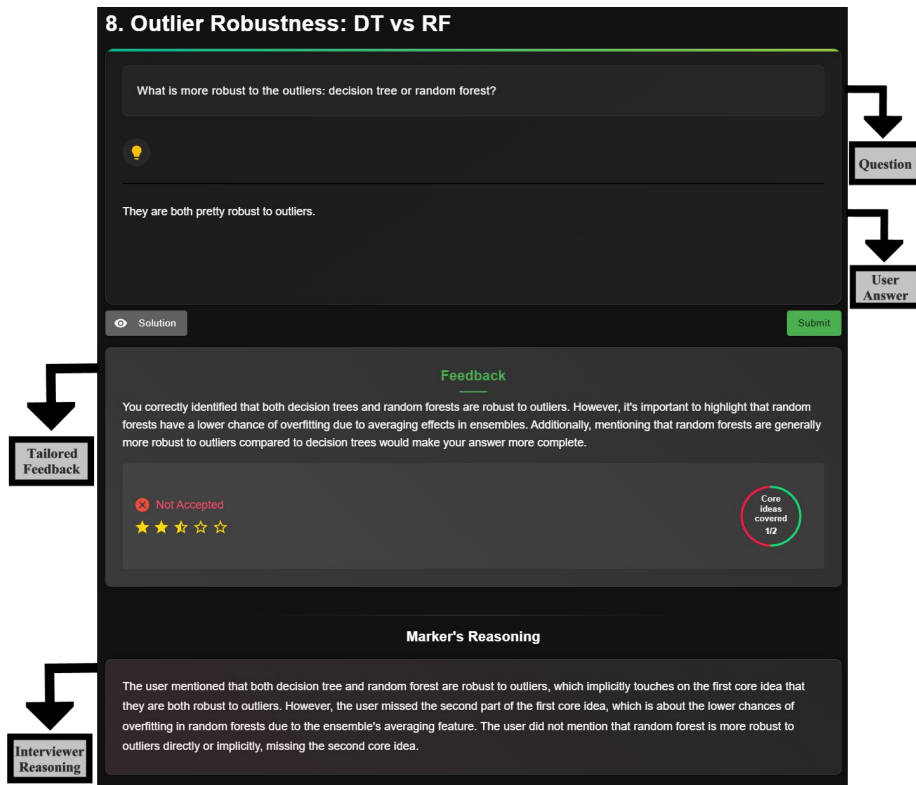
This is what we solve with NeuraPrep - a leetcode but for Data Science / ML roles.

It has over 400 questions asked at real interviews, together with coding exercises and quizzes!

The flagship: an autonomous technical **phone interview**, first in class.



# Text based interview prep



# Phone interview



# NeuraPrep – try it out for free!

Even though we got a growing paid subscriber community, we are still gathering feedback from people.

We are giving free credits to students in this virtual room in exchange for feedback!

Make an account at [neuraprep.com](https://neuraprep.com) and tell me your email you used to login or send it to:

[team@neuraprep.com](mailto:team@neuraprep.com) to add the free credits.



# Questions

Connect with me at:  
<https://www.linkedin.com/in/tidor/>



Don't forget to check out [neuraprep.com](https://neuraprep.com)!