

# CANDIDATE ASSESSMENT

Dear candidate,

✓ Thank you for taking the time to work on this technical assessment. It is part of our recruitment process for a position in Data Science.

The assessment is designed to showcase your coding skills, problem-solving skills, your ability to generate insight from data, and how you communicate your conclusions.


Please submit the following in no more than **two weeks**.

## A. Data science technical skills. Build a classification model choosing at least two of the problems below:

- ✓ 1. Classify facial expressions using the following dataset:  
<https://www.kaggle.com/datasets/aadityasinghal/facial-expression-dataset>
- ✓ 2. Classify Arabic poetry based on the poet's category using the following dataset:  
<https://www.kaggle.com/datasets/fahd09/arabic-poetry-dataset-478-2017>
- 3. Keyword Spotting (KWS) systems enable hands-free speech recognition experience by detecting a trigger phrase used to initiate the interaction with a device. You're challenged to build an accurate KWS with a lightweight classifier using the following dataset:  
<https://ai.googleblog.com/2017/08/launching-speech-commands-dataset.html>
- If you face any difficulties or find any hindering ambiguities, please feel free to send your questions via email.


You should at least cover the following in your solution:

- ✓ 1. Explore the data and use visualizations to explain to the reader the nature of the dataset (EDA).
- 2. Audit the data for missing or extreme values and deal with it reasonably, if there are any.
- 3. Mention the type of model that is suitable for the problem and describe your criteria for the selection (Decision tree, linear regression or deep learning, etc.).
- 4. Provide a report on the performance of the model and suggest recommendations for future improvements.


- 
5. Provide an error analysis section on the outputs of the model.
  6. Highlight any assumptions you made.
  7. Select a smaller dataset size if you encounter computational or memory issues.

These are the minimum required sections, please do not hesitate to add extra sections as you see fit. Higher model accuracy is not the most important criteria of the assessment; the process of developing the model is critical.

## B. Optional. Technical writing and design. Please choose one out of the two prompts below:

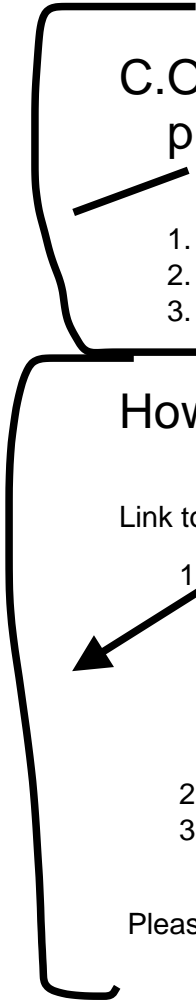
- 
1. Choose any research paper in the domain of deep learning and machine learning (or their applications).
    - a. Write a summary of a research paper of your choice.
    - b. Share with us why you selected this paper.
    - c. Share with us the limitations of the authors' work.
    - d. Comment on how one can benefit from the authors' work.
  2. (Purely hypothetical scenario) Write a detailed technical plan and solution design.

One of our clients is a famous YouTuber who occasionally hosts guests on their channel. Our client asked us to develop a suite of solutions to help them optimize the content. The client requires the following:

- 
- a. Real-time processing and storage of the live stream
  - b. Real-time transcription of the audio (the host and guests are Arabic speakers)
  - c. Real-time translation of the transcribed audio to three languages: English, French, and Chinese
  - d. Emotion recognition of all speakers
  - e. Logo recognition of any logo that appears in the live stream

Please write a technical plan on how to build and deploy a solution that meets the client's requirements. Please make sure you include the following:

- Any diagrams that demonstrate your chosen solution architecture
- Requirements from any other team (web developers, data engineers, IT), and hardware requirements
- Time estimates on the completion of the project and milestones
- Selected technologies (ex: software, middleware ...)
- Any questions you would like the client to provide answers to in order to improve your proposed architecture further
- Reference to the pre-trained models you propose to use or your plan for training a new model (or fine-tune) should there be a need

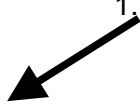


C.Optional. Personal. Please choose two out of the three prompts below:

1. What inspires you about your career? Optional: who inspires you in your domain?
2. Tell us about an unusual habit of yours or about a unique trait?
3. What's something outside of your work that you're passionate about?

## How to submit your work:

Link to a GitHub repo that includes all the below.

- 
1. Data science technical skills assessment:
    - a. Include Readme on how to install the required libraries or how to run your Dockerfile
    - b. Optional: Dockerfile to easily run your Jupyter Notebook
    - c. The final model
    - d. The solution (Jupyter Notebook)
  2. Optional: Answer in the form of a PDF to the prompts on technical writing and design
  3. Optional: Answer in the form of a PDF to the personal questions

Please feel free to contact us for any clarification. We wish you the best of luck.