The Alien Chat

# Regression Problem, 150pts

Aliens love to communicate in weird formats. This time they are mixing a set of signals that may look like noise to regenerate their real communication channel (an objective signal)

As a data scientist, you should identify the minimum set of input variables that aren’t noise to identify the signal the aliens use.

OBJECTIVE:

Define an algorithm to correctly recover the hidden signal. Use a limited number of variables for reconstruction.

FEATURES:

There are 25 features without descriptions.

TRAIN SET:

1,000,000 signals have been captured that include the hidden message (objective).

TEST SET:

There are 2,500 new observations for which the hidden signal needs reconstruction.

DELIVERABLES:

* Summary (must): Explain your approach to solve the problem to any hackathon supporter of the Data Science challenge.
* Regression: Share your regression results of the test set. Include only the ID, SIGNAL\_VALUE columns. You will get back your R^2 score\*

**\*** Your R^2 score represents the number of points you will obtain. Zero being 0 points and One being 100 points per signal! You can share your regression as many times as you wish, your highest value will be reported as the number of points obtained.