

Artificial intelligence

Coursework Document

Stephen anderson

n0786072

Contents

[General Explanation 3](#_Toc55939431)

[System Requirements 3](#_Toc55939432)

[Artificial Intelligence Techniques 3](#_Toc55939433)

[Program Explanation 4](#_Toc55939434)

[Chat Log 5](#_Toc55939435)

# General Explanation

My system is a chatbot written in Python using AIML and aspects of scikit-learn to allow for pattern and sentence matching and basic responses. The theme of my chatbot is about getting a UK motorcycle licence and the varying categories within it. The user can ask about the processes they need to go through for a licence, as well as asking about the different categories that are based on age and/or experience. It also has a list of different models of motorcycle based on style that the user can ask for a recommendation of. My goal is to help the user get a better understanding of their options as there are four licence categories with multiple paths to and through them. There are also other training and test certificates that are required and can do with some more explanation due to the complexities involved. My chatbot is also capable of letting a user select an image and request the identity of a vehicle in it.

# System Requirements

* Respond to a variety of simple greetings in an appropriate manner
* Respond to any small talk i.e. the user asking how the chatbot is doing
* Tell the user which licence(s) they are eligible for based on their age
* Explain the different tests and training the user would need to go through to legally be allowed to ride a motorcycle on UK roads
* Allows the user to use data requests to Wikipedia that return the most recent information
* Give a random recommendation for a motorcycle that fits a certain style to let the user look into them a bit more
* Allow the user to request the identity of a vehicle in an image

# Artificial Intelligence Techniques

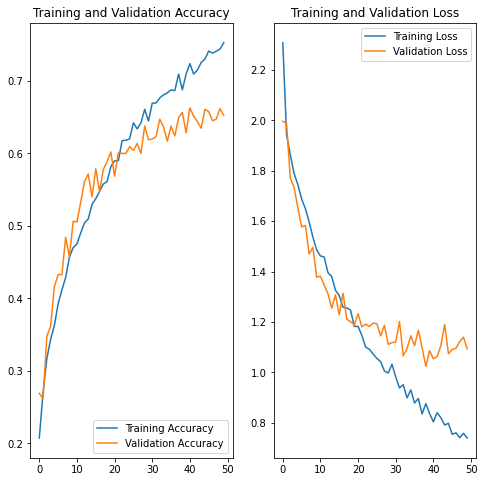
My chatbot uses two primary methods for determining the correct output based on the user’s message. First it uses the python AIML module to load and match the message to a predefined set of rules before determining if the TF-IDF model is required to find a specific response from a CSV file.

The artificial intelligence mark-up language file uses pattern matching, not too dissimilar to how regex works with characters, to decide its response. It then uses templates along with other tags that allow for some more complex behaviour to form a response that matches the user’s input most appropriately.

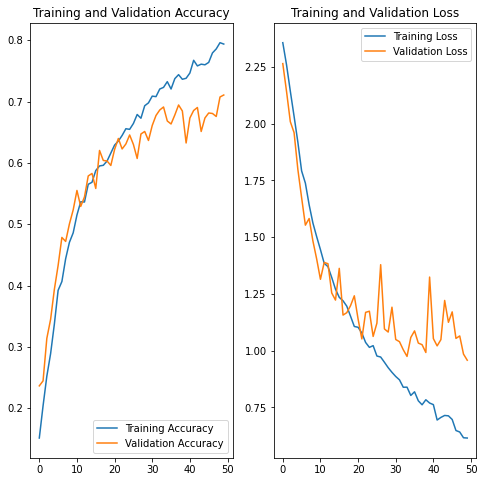
If the system has determined that the user wants to ask a specific question about some of the terminology or ask for a recommendation of a motorcycle, then it will use the TF-IDF model. This is where it calculates a list of the words in a response and gives them weighting based on their uniqueness in the sentence. This means that words that frequently show up in the text are not weighted as highly as if the bag of words model was being used and were weighted purely on frequency. This means that common words such as: “the”, “a”, “this” and “that” are generally not used as matching criteria. This TF-IDF value can then be used alongside calculating the cosine similarity to determine a match. This is done by calculating the vectors of two inputs before putting these into the cosine similarity equation and if the vectors are pointing in roughly the same direction – if the sentence is roughly the same – then it will tend towards 1. This allows the program to then see which value matches the closest and output the response. If no close match is found, then a different response can be given to the user.

The machine learning and classification model use a convolutional neural network (CNN) with 4 main layers and a 5th “Dense” layer that applies weights and processes all of the outputs from the previous layer in every node. There are some other processes being performed onto the data, both when loading it and when training the model. The images are normalised by being converted to grey scale to remove the colour from the images and therefore removes an extra variable from the model. As part of the training process, the model uses preprocessing to apply random factors of mirroring, rotation and zoom to the images to help reduce overfitting. The model is then trained and put through 100 epochs (iterations) using batch sizes of 32 images at a time to help speed the process up.

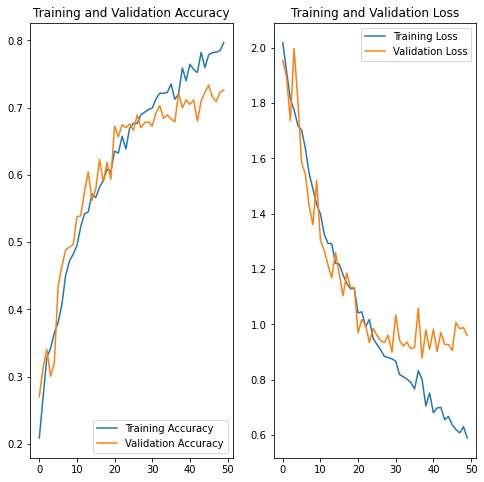
Each epoch is used to predict a class and compared to the actual image class, then producing a validation accuracy – how many correctly identified image classes there were. This data is then plotted on a graph using matplotlib.pyplot. I used multiple models for identifying the best fitting one as can be seen by these graphs:



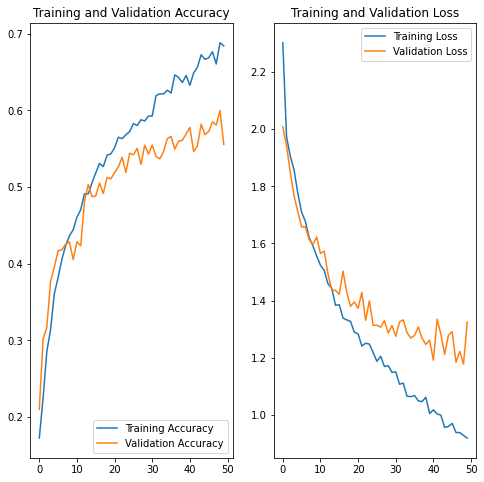
Figure



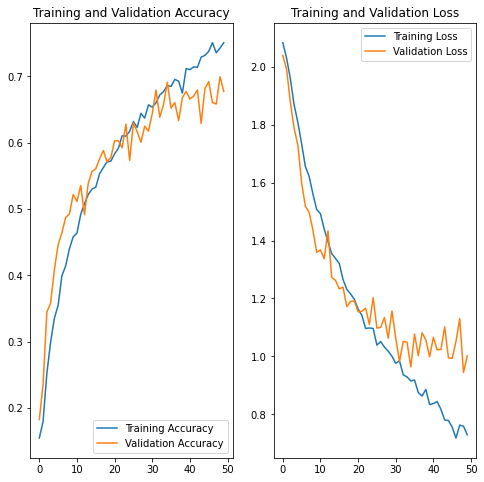
Figure



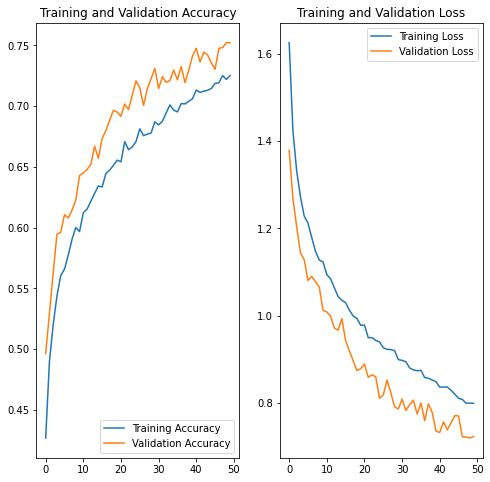
Figure



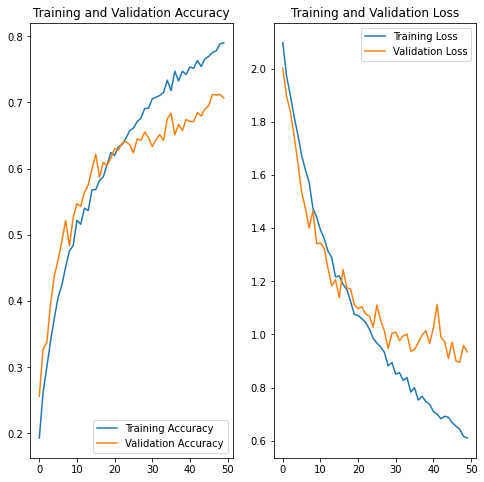
Figure



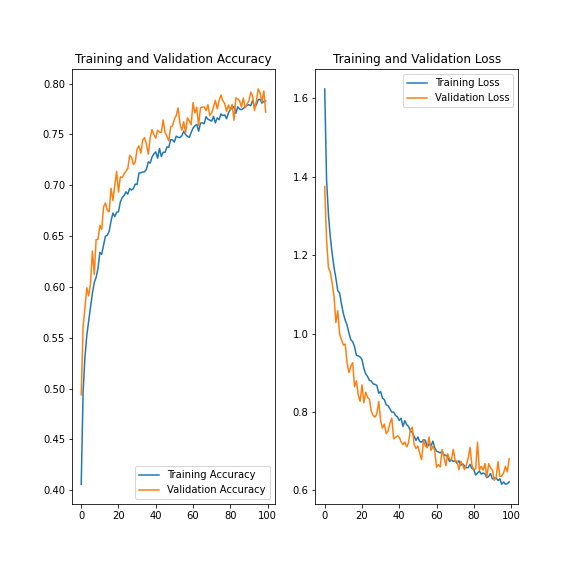
Figure



Figure



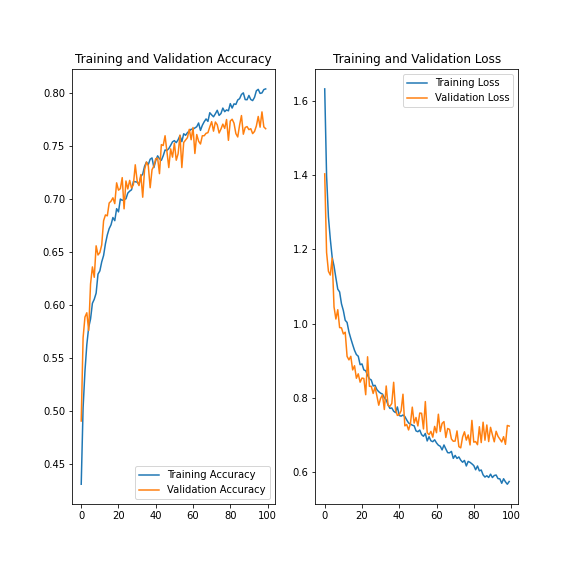
Figure



Figure



Figure



Figure

In each of the above figures, looking at the graph on the left side, the training and validation accuracy can be seen. If the orange validation accuracy line is consistently above the blue training accuracy one then the model is likely unsuitable due to “underfitting” where the model is likely not complex enough. This can be seen quite clearly in figures 6, 8 and 9. If the reverse is true and there is a large difference between the orange line and blue line where the blue one is above it then this is “overfitting” which occurs when the model starts to get too good at predicting the training data but not the dataset as a whole. The right hand side can also demonstrate underfitting or overfitting due to the loss; a higher orange line and lower blue line shows overfitting whilst the opposite is true for underfitting.

In the end I decided to use the model from Figure 8 as it has a good validation accuracy score with only a slight disparity between the training and validation accuracy. If I had used more epochs, I think I would have started grossly overfitting the model but around 90 is the sweet spot according to my graph – the two lines overlap each other at about this point. This model’s main features are that it uses an extra 128 node layer and the dense layer is 256 nodes; for my dataset it was the most effective and accurate that I tried.

# Program Explanation

There are a couple of imported modules, with ‘aiml’ and ‘sklearn’ being the only ones that needed to be installed separately. The ‘re’ module is short for regex and is used for data pre-processing. The two ‘sklearn’ modules are the ones used in the AI aspect of the chatbot.

The function ‘normalizeString’ converts the passed-in string to lowercase and then removes all non-alphanumeric characters except spaces, replacing them with a space. This is to help prevent issues with non-whitespace punctuation not being properly recognised either by the AIML or the TF-IDF processes.

The next function ‘fetchInfo’ is the function that is used to read from a .CSV file, match the line in the file that is most appropriate to the user’s input and then return it. It does this by first opening the file with the csv module then iterating through each row. For every row, it separates the first half of it – the question – into separate words using the in-built string manipulation of python. It then goes through the process of manipulating and matching the variables. If there is a match, then the row’s data aspect – the second half of the pair – is returned.

‘answerQuery’ is used as a wrapper that simplifies and modularises the program by allowing only a unique response to be output based on the result of the function ‘fetchInfo’ when looking at the Q/A pairs.

The last separate function is ‘getSuggestion’ and is similar to ‘answerQuery’ but is designed to recommend a random bike from a list of each style.

The next few lines are just initialising the AIML kernel to be ready for inputs and a greeting message that helps to explain a little bit about what the chatbot does.

The while loop will keep looping through until the user exits the program through the console or a keyboard interrupt such as Ctrl+C. Until then, it gets the user to input a message then uses the AIML kernel to determine the response before then processing that response.

The first part of the ‘if’ statement is related to what happens if an input is poorly handled and doesn’t return a proper response based on the default values. The ‘elif’ checks to see if the response is one of the special codes and if it isn’t then it just prints the response from the AIML kernel. If there is a special code, then it can do a particular response. If the code is ‘0’ then the program exits with a farewell message. If it is ‘1’ then the user has requested some information which will be searched for in the Q/A file. However, due to the extreme similarity of searching for the ‘A’ license category or the ‘AM’ licence category I needed to change the Q in the file to be for “moped” instead of “AM” and so the user input needs adjusting as such. Unfortunately, this extreme similarity can’t be solved due to the inaccuracy of the variable type losing such precision. If the code is ‘2’ then the user is looking for a bike recommendation based on a style. The input needs to be adjusted due to synonyms in the correct names of the style of bike. Finally, if the code is ‘-1’ then it means that the AIML did not recognise the input as any particular pattern and is resorting to the default value.

The user can choose to open an image for classification. This is done when the correct path through the program is taken and a tkinter filedialogbox object is created and called to allow for an interactive graphical way of selecting a file.

Once an image is selected it is loaded and converted to a type that can be used by my CNN model. The prediction can then be performed using the loaded model which will return a set of probabilities which are then converted into a single integer index that can be used to identify the class from an array of the class names.

The “Evaluate\_Prediction” function is used as the main wrapper and function call for the predictions and then handles that prediction before printing the result.

## Task 1 Feedback Improvements

I took the feedback from the first submission and added in the functionality for the Wikipedia Python module to make requests and also using the Python “requests” module to make custom API requests to the WikiMedia API so that I could fetch specific sections of a Wikipedia page. The exact request strings can be seen in the “QA.csv” file.

## Task 2 Feedback Improvements

## Task 3 Feedback Improvements

# Conclusion

# Chat Log

## Log 1

Loading Coursework-Bot.xml...done (0.00 seconds)

Kernel bootstrap completed in 0.00 seconds

Hello and welcome to this chat bot that can assist you in the processes of obtaining a motorcycle license in the UK. Please feel free to ask any questions that you may have. I can even suggest a few bikes for you!

> Hello

Hello! How are you?

> I am good thank you

I'm glad to hear that!

> How are you?

I am doing well, thank you for asking

> Do you know which licence I can get?

I don't know. How old are you?

> I am 20 years old

You are restricted to the A2 licence.

> What is the a2 licence?

The A2 licence allows for a motorcycle to produce power up to 35kW (about 47bhp) with a power to weight ratio of no more than 0.2kW/kg. There is no restriction on engine capacity but the bike must not be derived from a vehicle that has more than 70kW (about 94 bhp).

> What is the practical test

The practical test is broken down into the Module 1 test and the Module 2 test.

The Mod 1 is the off-road aspect of the practical test. The test takes roughly 20 minutes and includes manoeuvring the bike manually. It also features the manouvres such as the figure of 8, the U-turn and the slalom. You will also be asked to show the ability to ride slowly in a straight line and coming to a controlled stop after cornering. Finally, you will be asked to demonstrate the emergency stop and hazard avoidance after cornering. Before each of those manouevres you will go through a speed trap where you need to ride at a minimum speed of 30km/h (19mph) on a moped or 50km/h (31mph) on a motorcycle.

The mod 2 is comprised of two main sections. Before you're allowed on the road, your eyesight will be checked by reading a number plate from a distance of 20m. Then you will go through the 'show me, tell me' questions about the vehicle you will be riding. This will cover things such as brake controls, minimum tyre tread depth and some basic maintenance like checking the oil level.

You will then be taken out onto the road where you will be asked to perform normal stops, angled starts (like pulling out from behind a vehicle) and sloped starts (such as on a hill.) You will be receiving directions via radio from your examiner who will usually be following on a motorcycle. After this section is over you will be asked to independently ride, following some basic directions. You will not have access to a sat nav/gps device and you will not fail if you go off the route, as long as it is safe; your examiner will give you directions to get back on track.

For both tests you will be given your results once you are back inside the test centre.

> What is the CBT?

The Compulsory Basic Training (CBT) is about a day-long course where an instructor will teach you the basics of riding a motorcycle on UK roads. You will also be tested to see if your eyesight meets the legal requirement, usually by reading a number plate at 20m.

As your training will require riding a motorcycle, you will need appropriate clothing and equipment. Some schools provide more clothing than others so please check when you book your training. Your day will start with understanding the controls of the motorcycle/moped you will be doing your training on. This will then move to being able to manually move the vehicle around, such as taking it on and off its stand(s) and wheeling it around a course in a controlled manner before coming to a stop.

Next will come the on-site riding which will cover travelling slowly, riding in a straight line before coming to a controlled stop, coming to an emergency stop, riding in a figure of eight, riding in a u-turn and simulated left and right hand turns at junctions. Before you are allowed on the road, you will be coached on the legal requirements of you as the rider and also your vehicle, the responsibility that you have and best ways to ride safely.

Once you have done all of this you are allowed onto the road with one instructor per two students. The on-road practical session will be a 2 hour long ride on varying types of roads for at least two hours where you will be expected to demonstrate the skills that you have learned from the on-site practical training such as the u-turn and how to safely turn left and right at junctions. It is also expected that you will be taught how to correctly go around roundabouts in this session.

> Suggest a good sports bike

Ducati Panigale V4R

> Recommend a naked bike

KTM 790 Duke

> What is the AM licence?

The A licence has no restrictions on engine power or capacity. It also allows for the use of a trike regardless of its power or a sidecar to be attached to a motorcycle.

> Thank you

You're welcome

> Goodbye

Bye! It was nice talking to you. Don't forget your lifesavers ;)

## Log 2

Loading std-startup.xml...done (0.00 seconds)

Loading Required-Files/aiml\_1.aiml...done (0.00 seconds)

Loading Required-Files/aiml\_2.aiml...done (0.00 seconds)

Loading brain from bot\_brain.brn...done (33 categories in 0.00 seconds)

Kernel bootstrap completed in 0.00 seconds

Hello and welcome to this chat bot that can assist you in the processes of obtaining a motorcycle license in the UK. Please feel free to ask any questions that you may have. I can even suggest a few bikes for you!

> what is the cbt

In the United Kingdom, the term Compulsory Basic Training (abbreviated to CBT) is a preliminary vehicular training course which must be completed by people wishing to ride a motorcycle or moped unaccompanied on the road, and remains valid for 2 years upon completion. It was introduced in Great Britain on 1 December 1990 as a means of reducing accidents on the road caused by inexperienced drivers by reviewing aspects of riding both on and off the road with a qualified motorcycle instructor registered with an Approved Training Body (ATB). If a full car licence was obtained before 1 February 2001 it is not necessary to complete a CBT course to ride a moped.A CBT allows a rider to ride a moped up to 50 cc with a provisional licence from 16 years of age, and a motorbike up to 125 cc from 17 years of age.Compulsory Basic Training consists of five elements;

Element A - Introduction and eyesight test.

Element B - Learning the controls of the motorcycle.

Element C - Off-road riding.

Element D - Road briefing and safety talk.

Element E - Practical on-road riding.Northern Ireland introduced CBT in 2011. As of 21 February 2011 learner riders in Northern Ireland must complete CBT before being allowed on the road unaccompanied. Those who obtained their provisional motorcycle/moped licence before the implementation of CBT in Northern Ireland had until 21 February 2012 to pass their motorbike test, or else take CBT to continue riding as a learner. Those who passed a car test prior to 21 February 2011 have their full moped entitlement preserved without having to take CBT.

> what is the mod 2

== Module 2: on-road riding ==

The second module is the on-road test, but also includes an eyesight test and safety questions as well as road riding.

=== Eyesight test ===

The candidate must read a number plate on a parked vehicle 20m away. Failure at this point means the test is cancelled.

=== Safety questions ===

The examiner will ask the candidate several safety questions (there is no prescribed number). A full list of the safety questions is available [https://www.gov.uk/government/publications/motorcycle-show-me-tell-me-vehicle-safety-questions/motorcycle-show-me-tell-me-vehicle-safety-questions here].

=== Road riding ===

The candidate will be in contact with the examiner using a radio and must complete normal stops, angle starts (pulling out from behind a vehicle) and a hill start.

=== Independent riding ===

The candidate is given an approximate location to navigate to and the examiner follows, observing the ride. The candidate is not tested on whether he or she arrives at the intended destination, only the quality of riding. The candidate must follow the traffic signs and road rules.

If the candidate passes, the examiner will give instructions on how to change the provisional licence to a full one for that class of motorcycle.

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/Bicycle-3.jpg

That is a Bicycle

> what vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/Motorcycle-2.jpg

That is a Motorcycle

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/Car-1.jpg

That is a Car

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/Helicopter-3.jpg

That is a Helicopter

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/My-Images/FB\_IMG\_1535013746022.jpg

That is a Motorcycle

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/My-Images/IMG\_20190826\_125944.jpg

That is a Bus

> which vehicle this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/My-Images/588421.jpg

That is a Helicopter

> which vehicle is this

Loaded: /home/steppy/Documents/Year-3/Artificial Intelligence/Coursework/Test-Images/My-Images/20180813021954\_1.jpg

That is a Motorcycle

>

Goodbye!

## Log 3

Loading std-startup.xml...done (0.00 seconds)

Loading Required-Files/aiml\_1.xml...done (0.00 seconds)

Loading Required-Files/aiml\_2.xml...done (0.00 seconds)

Loading Required-Files/aiml\_3.xml...done (0.00 seconds)

Loading brain from bot\_brain.brn...done (39 categories in 0.00 seconds)

Kernel bootstrap completed in 0.00 seconds

Hello and welcome to this chat bot that can assist you in the processes of obtaining a motorcycle license in the UK. Please feel free to ask any questions that you may have. I can even suggest a few bikes for you!

> check that softail rocker c is chopper

That's true.

> check that softail rocker c is cruiser

That's true.

> check that softail rocker c is motorcycle

That's true.

> check that sprint 125 is vespa

That's true.

> check that sprint 125 is motorcycle

I'm afraid that's not true.

> check that sprint 125 is moped

That's true.

> check that kawasaki make kx450

That's true.

> i know that gs500e is naked

OK, I will remember that! Thank you.

> check that gs500e is naked

That's true.

> check that gs500e is suzuki

Unfortunately I don't know. It could be true or false

> i know that sprint 125 is motorcycle

I am sorry but that contradicts my knowledgebase!

> i know that gs500e is naked

I already know that! Thanks anyway

> bye

Bye! It was nice talking to you. Don't forget your lifesavers ;)

## Video Links:

<https://youtu.be/PShU2eic1FY>

<https://youtu.be/91Z-fw5L16M>

<https://youtu.be/ZPLGWiS7oLk>