Reese Jones' Writing 1 Corrected:

Our project aims to contribute to the pool of computer tools which allow scientists to do their jobs more efficiently. We are proposing a system that would enable biologists studying moths to have a more efficient and streamlined process for collecting and understanding data. Currently biologists must gather and sort data entirely by hand, which presents significant limitations in that one person or team of people can only process a limited amount of data. We aim to assist in that field by creating a two part system of tools which streamlines the processes of collection, classification, and sorting of data regarding moths. The two portions to our project are as follows: a mobile application interface for submitting images of moths to the project, and a machine learning algorithm trained to identify and measure moths of different species. The completion of this project will enable biologists and others researching moth populations to spend more time focusing on the actual implications of the data and less time on processing large banks of raw information by hand.

Our project aims to help significantly decrease the number of working hours spent on processing and data that is collected in the field. We aim to do this by creating a mobile application which will allow users to submit photos for classification without any other work or steps involved. The mobile application will create a system whereby the submission of data is simple and as streamlined as possible. In addition to simplifying submission of images, the mobile application will also properly format data in order to allow the whole system to function more efficiently. Currently, a simple system for data submission and preprocessing simply does not yet exist in the field and, when completed, this system will greatly expand the pool of potential data available to biologists through crowdsourcing of data points. In addition to the mobile application, the project will also involve a machine learning algorithm which will be calibrated through multiple rounds of testing and reinforcement using previously obtained data in order to estimate which species of moth is present in an image. The process of machine learning will take place using data available and already understood, so as to ensure accuracy with species identified previously by biologists. Through the use of such an algorithm, our project will enable researchers in the field to track and identify certain species of moths through a simple photo, rather than having to painstakingly collect data in the field and analyze in a lab at a later point. This will allow for professionals in this field to be exposed to more data points, thus improving the accuracy of their research and predictions. This project's main objective, achieved through the tools described, is to streamline the efficiency and availability of systems for identification of moth species for biologists actively doing research in this field.

The completion of this project will permit the biologist community at large to more effectively research patterns in moth populations, which is becoming an increasingly important issue as the world deals with the onset of climate change. Through both a mobile application and a web based machine learning algorithm, our project will provide the utility of computers to the scientific community in an area where it is much needed. The sensitivity of moths ecologically has led them to be called an "indicator species" in some instances, which speaks to how important it is that biologists be able to monitor their populations across the world. This project will ensure that the evaluation of moth populations is a simpler process that is more accessible to researchers across the globe. Our project seeks to enable researchers to more effectively monitor the health of our environment. Instead of forcing people to complete a job that computers are much more suited for, our project will allow researchers to focus more on the substance of the data rather than its collection and processing.

Ben Giangrasso's Writing 1 Corrected:

Advancements made in computer-aided research are fundamental to scientific growth by allowing researchers to gather information more efficiently. In biology, scientists aspire to classify and catalogue the millions of species that exist on Earth. Species classification is necessary to the biological community because it allows researchers to study differences between different organisms, allowing them to learn more about their effects on their environment and other species. Through machine learning, our project will allow scientists to classify species of moths through the straightforward user interface of a website and mobile application. Moths have a critical ecological role as both pollinators and as prey to a variety of other species. The classification of their species is important to understanding the role that they have in their respective environments and impact on other forms of life. A simple interface will streamline moth classification by swiftly informing researchers of a moth's species given its image. Our user-friendly application will reduce the time and manpower required for moth classification by transitioning research from the field to a virtual platform.

We aim to develop an application that allows users to submit pictures of moths from their camera or photo gallery where the image will be processed by our two machine learning models. Our first model will be used to detect moths in the users' submissions. Then, the second model will classify the species of moth to the user. Both of our machine learning platforms will initially be trained through a large dataset of moth images. We intend to add the moth images submitted to our application into our dataset, allowing our machine learning models to continuously develop and become more accurate. We plan to uphold our goal of providing quick results by performing image classification calculations on a cloud platform. The website and mobile application will simply be a vehicle for users to submit and receive data from our cloud. Cloud computing will provide more efficient processing power so that our applications can focus on giving a faster interface when returning information back to the user.

Aside from our machine learning image classification, our application will give moth researchers the option to view image submissions to personally classify them. These classifications will improve the accuracy of our image processing by validating each image that is submitted. This will benefit users by supplying them with a platform to view different species of moths from our dataset while being able to classify the moths through a snapshot of their cell phone camera. Our goal is to create a mobile platform that will remove the time needed to conduct physical field research to classify moths. Instead, researchers will be able to use machine learning to get their results in seconds.

Our project will prove to be useful on both a practical and scientific level. Users will be able to photograph and understand a moth's species to further determine the effects that they play on their environment. Users can use this information to determine the types of plants that exist in the environment from the species of moth and the plants that they pollinate. Similarly, users can also determine the animals that exist in the same environment by deriving the group of predators that feed on a certain moth. Our project will also give moth researchers a location to view pre-existing images of moths with their respective species classification to conduct research from a screen instead of in a laboratory or a moth's environment. Every user of the application will be benefiting the study of moth species by adding each of their pictures to our dataset. This will give researchers the opportunity to study moth species outside of the field. The image classification step in their research will be completely automated, leaving them to make further contributions to analyzing moth species without having to spend time observing the physical characteristics of each moth.

Abid Ahmed's Writing 1 Corrected:

One of the greatest benefits technology has brought to society is automation. The ability to make a task less labor intensive, less time consuming, and easier is invaluable. It allows a person or company to be more productive. In some cases, technology can find information humans are unable to detect. Another benefit of technology is that it creates uniformity. When humans are tasked with analyzing data, the information they collect might be messy and inconsistent. Websites and mobile apps allow people to view data in a consistent and clean manner. Thus, the goal of our project is to bring these technological advancements to a group of researchers that specialize in studying moths. They are in need of an easy to use tool that classifies the species and size of a moth in a picture.

Researching and analyzing moths answer many important questions regarding our environment and climate. First, moths indicate the health of an environment and ecosystem. According to The Guardian, moths pollinate flowers like bees do, and they make up most of the diet for certain birds and bats. According to The Science Times, researchers have noticed moths are getting smaller and migrating away from tropical environments due to the increase in global temperatures. Our application will be able to identify moths in different or new areas to determine the migration of moths. Scaling the application will expand the reach of data collection and moth identification. Thus, we believe that it is worthwhile to invest our time in building this technology for moth researchers, and we believe that it is worthwhile to invest your money in this cause due to the environmental impacts moth research has.

To achieve our goal of building this project, we have three major components that we will build. First is the user interface for our project. Our plan is to use a mobile application to provide users with an easy to use and intuitive interface. We want users to contribute to moth research, so we will allow users to take pictures of moths and have them be classified by our software. This reduces the labor requirement of the biologists. Our mobile application will allow researchers to have access to the same functionality as normal users, and we will also allow them to view and filter the crowd-sourced images of moths. They will also be allowed to manually classify moths due to their expertise. Crowd-sourced images of moths once the product goes live will allow moth researchers to have a free and continuing source of information.

Our final two components of this project will be creating artificial intelligence to automate the detection of moths in a picture and the identification of the species and the software that communicates with the mobile application. Our goal is to teach the artificial intelligence so it becomes capable of completing the tasks listed above while being able to generate an outcome almost instantly. By creating an artificial intelligence, we will be able to successfully automate the classification of moths and save researchers a significant amount of time since researchers such as Dr. John Pickering describe the process of classifying moths as laborious and time-consuming. In addition, we will need to create a program on the server that allows for communication between the mobile application and our artificial intelligence. We will also need a medium to store data and process user actions. Thus, our server will be key in integrating our project together while also providing timely and continuing support to our users.

This project will greatly reduce the amount of labor hours researchers will spend manually classifying moths. It will also allow average people to contribute to moth research and it will provide researchers with a free and large source of information. However, a significant impact is that this project can be applied to any field requiring the automation of data collection and classification. For example, this project can be applied to the collection of car data for law enforcement purposes. The possibilities with the results of these projects are endless and can be used for numerous purposes, not just moth research.

Writing 2 Corrected:

The use of machine learning for image classification is becoming increasingly relevant to the success of businesses. Google uses image classification to allow users to search for images through the submission of a related image. Ebay hosts a similar platform that grants customers the ability to search for products with a picture. This helps customers to find their desired items without having to guess the product name. Likewise, Pinterest has image classification software that can pinpoint specific items in a photograph, informing the user about the specific product and brand. Stanford University uses similar software for medical research and to detect lung cancer. Although we intend to use machine learning to recognize images like these companies, our primary goal is to classify preexisting images of moths that are not yet classified. Ecologists collect thousands of images of different species, but a majority of the pictures have not been analyzed. Our usage of image classification will withdraw the need for ecologists to step through each image by hand. Since we are not only classifying new pictures, our project will significantly impact ecological studies by linking scientists' research with an efficient, automated platform. Our team will primarily classify images from the Discover Life database. Discover Life is a growing compilation of information of around 1.4 million species. Discover Life currently classifies moths by requesting that the owner of the submitted pictures tags the time and location of each picture submitted. Then, an authorized user sorts the new pictures into buckets based on their shape, color, and pattern. Designated experts then look through each bucket to match each image to a species. Discover Life mentions that their human-based version of moth classification takes 35 seconds to classify each individual image. As large numbers of moth photographs are submitted to the Discover Life database, more time is required by researchers to personally categorize these images. Our trained machine learning model will replace this process with a trained algorithm that could classify these images at a fraction of a second.

The main audience for our product is biologists focusing their research on moths across the world. More specifically, the project will be utilized by Dr. John Pickering and other biologists and researchers who are contributing to the Discover Life project. Due to the nature of our project being centered around streamlining the Discover Life project's data collection, our completed product will present a useful data analysis tool for the whole moth research community. This product is focused on improving the efficiency of the research efforts at Discover Life, and being that no other comparable solution exists to satisfy their needs our product will serve to increase the overall efficiency of the Discover Life project. Upon successful completion of this project, the Discover Life community at large will have a new tool with which to analyze and identify moths quickly at scale. Additionally, being that the usage of our product will only require the downloading of a mobile application, it will be an approachable and easy to use utility which is available at no cost to the individual user. On the whole the business model of the project is simple and cost effective because once development is completed, all of the project infrastructure will be transferred to already existing Discover Life servers and will therefore incur no greater cost than Discover Life already has. For both of those reasons, the product will become a clear and obvious solution to the problems facing the Discover Life project in terms of collecting mass amounts of data.

In order to generate interest for our product within the biologist community and persuade researchers to use it, we will focus on disseminating information about the benefits that our product can provide to a research group. There are a number of reasons that our product will be easily marketable to the biologist community, the strongest of which being how simple it will be to utilize our system. For the community of Discover Life biologists who are focused on specifically contributing to the Discover Life project, our product will drastically reduce the human hours spent classifying data. That is a tangible statistic that is easy to convey to researchers through a demonstration of how fast our system will be able to go from submission of an image to the output of a classification as opposed to a human completing the same task. For those biologists and researchers who are outside of Dr. Pickering's project, our product will expose a wealth of information regarding moths and provide the ability to classify their own data using knowledge gained from the classification of the Discover Life data. The team's most effective tool for promoting the usage of our product will be the assertion that it will streamline the process of doing research about moths across the globe. In addition to that, the data collected through our product will enable future biologists and researchers to start into research with less of an initial investment in research, which will benefit the field by allowing more minds to contribute to the same research area. For both of those reasons, our product will be able to be spread about the Discover Life biologist community and accumulate users with ease.

As a result of improving the efficiency of the moth researchers, they will be able to make scientific breakthroughs regarding moths faster. Although many people see moths as pests, they are a very important species when it comes to the health of our planet. First of all, moths, like bees, contribute to spreading pollen. According to an article published by The Guardian, moths are second to bees when it comes to pollination, and moths are active pollinators in every continent except Antarctica. Secondly, moths are an important part of the food chain. According to the same article by The Guardian, moths are the main source of food for certain bird and bat species. In addition, certain moth species are active in the winter which allows birds to find their much needed sustenance in a time where food is scarce. Due to these characteristics of moths, moths are great indicators of the health of an environment and ecosystem. In addition to being environmental indicators, studying moths allows researchers to answer questions about climate change, a real and important issue in today's society. According to The Science Times, moths are getting smaller due to increasing worldwide temperatures. A reduction in size will lead to less eggs being laid by moths which will lead to less moths in the environment which does not bode well for pollination and feeding other species. In addition, certain moth species are migrating away from tropical climates due to the increase in temperature. As a result, classifying a moth, figuring out its size and location will allow researchers to identify the rate of climate change, changes in the worldwide climate, and the impact of climate change on other species.

Once our product is used widely, the data collected by the system will grow and allow for more insights and discovering trends. More moths will be classified, and a more diverse set of moths will be classified if the product is used worldwide. Researchers will not have to travel to different places to collect moth data and instead can use the data collected by other researchers. This will allow researchers to get more information faster which allows them to answer the pressing questions of climate change and the environment faster and more accurately. Due to the increased information and research potential, it will allow decision makers to make more informed decisions about environmental and climate change policies. However with the positives comes the negatives. One issue we foresee with scaling our product is environmental disturbance. Although researchers have been taking pictures of moths, there might be issues if we decided to crowdsource moth data by opening our product to the public and allowing them to submit pictures of moths to the application. It will help researchers in getting more data, but there will be regulation required if environments get disturbed with less experienced and trained people taking pictures of moths. However, we see the crowdsourcing aspect as a net gain due to the increase in information.

Writing 3:

[Link to our project timeline on GitKraken (Hover over items for detailed description)]

To build our system, we plan on implementing novel ideas by leveraging existing technologies such as Flutter for mobile application development, the Django REST framework for rapid development of middleware, and Google's Tensorflow machine learning framework. Although there are many services and machine learning models like Apple's Core ML framework, Google's ML Kit, ResNet50, and VGG-16 that offer image classification and classify animals, they are not tailored enough to classify and distinguish between animal species, especially moth species. Thus, the novelty of our system lies in the tailored analysis of images to distinguish between different moth species. To accomplish this we are going to train our machine learning models with a large and varied dataset of moth images so that it can detect the fine details that distinguish moth species from each other. In addition to the novelty of the species classification, we are leveraging Flutter and the Django REST framework to create a custom tailored mobile application for our customer, Discover Life researchers. To ensure the accuracy and confidence of our system, we are giving these professionals the ability to certify and add their own classifications to our automated ones. Many machine learning products are standalone in the sense that they do not utilize user input and just give an output. Thus, we are going against that idea and allowing professionals to give their own classifications if there is doubt regarding an automated classification. This will improve the overall health of our system and instill confidence since it is being checked by subject matter experts. Providing a custom tailored solution for image classification and adding the ability to certify predictions is what will separate our product from other various offerings.

Our design is broken down into three separate components: the machine learning platform that takes in image data to retrain itself and process unclassified images of moths, a user-interface that allows users to easily submit unclassified images of moths and receive classifications from the machine learning model, and an application programming interface (API) that is used as a medium to send data to and from the other two components. Each one will be built from existing, well-trusted technologies. First, the machine learning component will use Tensorflow, an open-source coding library that will allow us to build and tune our machine learning algorithms with ease. The machine learning code will run on the Colonial One Cluster, a platform given by The George Washington University that will accelerate the time needed to compile and run large computations. This will help us test our machine learning models at a faster rate, allowing us to spend most of our time reconfiguring our machine learning algorithm instead of waiting for our model to finish training. Our user-interface component will be built with Flutter, a coding interface that is used to produce high-quality mobile applications. This software's main benefit to our design is that it will help us produce mobile applications that are consistent across both Android and iOS platforms. The application will use Google Firebase, a cloud database that can save data on the cloud, allowing users to have an easier experience using the application. For instance, it will be used to remember login information of each user with an account so that users do not need to sign in to the application every time it is used. Google Firebase also displays application analytics. We will use these analytics to trace faults in our user-interface such as picture submission errors so that we can easily find errors while our application is being used. Our final component, the project's API, will be built with the Django REST Framework. This tool is used to organize our database of moth images and metadata in patterns that are easy for the user-interface and machine learning components to understand. This is beneficial for our design because it mitigates the amount of work needed to rearrange data when being sent across our three components. We are confident that we will be able to use these technologies effectively because they are each built to be flexible across a variety of uses. These tools are each made to help programmers build intricate projects by eliminating complexities that would have been encountered without their assistance.

The development of our project comes with relatively low overhead costs, due to the nature of the deliverables of our design. In terms of hardware, the only components which we are figuring into cost is the cost of running a webserver to serve the Django API which serves as the middleware between the Flutter application and the machine learning algorithm. The server that we are running currently only runs $15 a month based on our use case, but should demand be significantly higher in the future there are ways to scale our server or utilize multiple servers and load balancing with the only additional cost being more server space. For the Flutter application, the Firebase databases we're using for authentication and storage both have a very generous free tier which will allow us to use the services without cost for the foreseeable future. We will only be charged if our active user count exceeds ten thousand per month, which we believe to be unlikely. There are almost no costs associated with running the machine learning process for our team because in the development stage we are running training on the Colonial One Cluster provided free of charge, and in production the processing will be offloaded to the cluster of servers that DiscoverLife already has for their purposes. For each module in the project there is a different estimate for the total physical work - in terms of lines of code written - which will end up in the final version of our product. For the Flutter app the number is considerably higher than that of our machine learning module. Specifically, as a result of the large amount of boiler plate code and other files which are used to make up the building and packaging of two mobile applications, the estimate for the flutter application is somewhere in the neighborhood of six or seven thousand lines of code. This will include the boiler plate code required to generate screens as well as all of the additional services, libraries, shared utilities, and tests which will be built for the app. On the other hand, due to the nature of machine learning using Tensorflow, the estimation for our machine learning module is that the codebase will not exceed around 300 to 500 lines of code. This is due to the fact that the adjustments being made over time do not actually add code to the initial boilerplate, but just adjust the parameters being used. Lastly, the Django REST api we estimate will contain a codebase of about 2000 lines of code. This is because each endpoint and the related code to process and deal with data is about 200 lines for each endpoint. Overall that places us with an estimated overhead of around 10,000 lines for the whole project at the time of completion.

There are three major milestones (aside from total project completion) when it comes to project execution for our team. First, a point where each individual module can mock the flow of data from submission from user to processing and output with consistent formatting which will be compatible with other modules. For us this point should be reached before the end of November, as it is critical for continuing development and later allowing smooth integration. Second is the point where all of the module data flows are integrated in such a way that a photo submitted in the mobile app will trigger the proper flow of data such that the image ends up in the Django API store and that the machine learning algorithm can run a prediction on the image successfully and return that to the user. The timeline for this milestone is to complete this by the beginning of January, before we start the spring semester. The third milestone is actually having the Flutter app published to the App Store and Google Play store. The timeline goal for this is to have the first beta version sent out for approval by the beginning of March as the timeframe for approval varies widely.