

Student-At-Risk Predictor: Team Contract & Design Proposal

Motivation and Ideas:

Jason (Project Manager): As someone who has been at risk of failing a few classes, I could see a tool like this helping me immensely. Admittedly, I have also had to retake a couple classes due to a lack of understanding of material, as well as poor study habits and procrastination. I am confident that this tool will be able to support students facing similar issues, because the issue of studying or preparing for a course is usually not studying itself but studying the right material.

Instead of doing a brute force and studying over concepts and processes a student is already comfortable with, they can target their academic blind spots saving them time and increasing the efficacy of their studies. Moreover, the teaching styles of certain professors don't mesh well with every single student, and this tool can. If we can make this concept work for the purposes of Canvas and instructor support, then I imagine that the general workflow can apply to other areas of education and self-development (think engineering bootcamps, LeetCode/interview grind, learning modern technologies).

I am also looking forward to utilizing my professional experience in software development, project management, and generative artificial intelligence to further these personal goals and prove my skill as a technologist and leader. I want to grow in areas of project management that typically require soft skills such as team communication, time management, and resource allocation. I have spent most of my time in professional and academic teams as a developer, and I feel like I can use this experience to relate to and functionally support my fellow teammates.

My main technical idea for this project is training a GPT model for tailored academic content, something I created a [rough draft](#) for (further details in Artificial Intelligence section). It will provide our YouTube content recommendations in JSON format but requires further testing for programmatic access and more test cases (content topics). I have leveraged OpenAI integrations in my previous summer internship at Morgan Stanley and have been fascinated by how we can leverage the chatbot programmatically. I anticipate that it will continue to dominate the tech industry as a powerful tool and custom tooling for it will absolutely become an indispensable skill. Using it for educational purposes is a novel idea that I expect the team to capitalize on to solve a real world problem. While many students might use this functionality to cheat on

homework, I am a firm believer that it can be utilized as a force for academic growth and development.

In addition to leveraging generative AI, the team is also using the more traditional AI approach of regression to predict whether students are at risk. My academic interests lie primarily in data science and artificial intelligence. I took the course Algorithms for Machine Learning a couple of years ago, and I am currently taking Artificial Intelligence to support this project alongside our AI/ML lead. I recently made an image classifier using a convolutional neural network to distinguish between happy and sad images, and I think that I will be able to support the development of a regression model to predict at-risk student behaviors and outcomes. A conversation with our sponsor will be valuable to our understanding of how to approach the model, as he has worked with mining data from Canvas, and we will leverage training data of former “at-risk” students to predict that status in real-time.

From the perspective of a software engineer, there is much to learn from any senior design project. However, implementing APIs that interface with a custom GPT model as well as integrating a machine learning pipeline will teach me more about modern practices and initiatives across the industry, and how to build large-scale solutions from the ground up. I know I have fundamental knowledge gaps in system design and deployment, and I want to head some of those directives in the second semester.

I believe my role as a (technical) project manager will serve myself and my team well. My industry experience will be applied in this project, and I am excited to work with the team to meet our goals for both semesters. I am already getting to be a jack of all trades when it comes to software development, enabling communication amongst the team while also assisting in making technical decisions and tradeoffs across the stack as it begins to form. I am looking forward to learning more from the team and sponsor on how we can cultivate personalized solutions to academic challenges, and how I can become an efficient and dependable project manager along the way.

Justin (Frontend): Creating and designing a tool that is going to help monitor the progress and potential failure of a student is an incredible opportunity. Integrations such as this risk predictor

are going to help save students from drowning in their grades and is something I could have used at some points in my early academic career. This project also serves as a template to widen my template when it comes to front-end design outside of a typical webpage. Integrating this app into the Canvas Infrastructure LTI is going to pose as a new challenge and will allow me to explore new frameworks that could cut down on development time. Python frameworks are what come to mind since the application will have a machine learning mechanic and frameworks such as Taipy and Flask could prove useful along with their extensive documentation. The graphical representation of a student who is at risk will be up to the team's discretion and should be a representation that is easy to understand and does not cause the reader to question the assessment. The goal in mind is to have a minimalistic application that the user will see naturally. The journey would be the instructor would sign into their instructor portal and would navigate to a course of their choosing. Once they have loaded onto their course page, there would be a panel describing what students are at risk of failing the course and **how at risk** they are. The alternative view would be a meter describing the student's risk when an instructor goes to grade an assignment for a student.

Revisiting frameworks to use for this project, Canvas is built on Ruby through Ruby on Rails. While this is a foreign backend language, for my needs I believe I can utilize basic JavaScript and potentially a React framework. Since the application will be based in the website itself, there will not be much room to create my own app since I'll be limited in terms of the UI. The remainder of the semester leaves much time to find implementations that can serve me best and grants me the ability to test and debug as well as write extensive documentation.

Moving away from the technical aspects of the project, I am excited and nervous to be able to learn how to implement frontend projects into already existing applications. My time at the University of Central Florida has taught me some of the basics of web development, however much of my knowledge has been from outside the lecture hall. This will also give me the opportunity to further learn and understand the intangibles of front-end development. Features such as the certain coloring for risk assessment, the location of the tool in the UI, the users experience in general are something I have little experience in, and I am hoping this project will help me mature in that area.

Teamwork within a project is always something that needs to be prioritized. Working on a production level project will emphasize that, and I am excited to collaborate with a cross-

functional team and our sponsor who in my eyes is acting as our client who we will be delivering to.

University coursework will always serve to practice our skills and fundamentals and demonstrate that to the professor by submitting our work and that will normally be on our own account outside of a project. However, delivering a product to a client will require me to take all elements of everything I've learned in my time here and apply them to this project.

[David Santamaria, AI/ML interface]: Having been an ‘at-risk’ student myself, I’ve experienced the stress and anxiety that comes with high expectations to do well in the next task to save a grade. I’ve seen smart students that perform exceptionally outside of the classroom environment in extracurriculars go through depressive bouts because of the fear of having to use grade forgiveness and repeating a semester, or having a mark on their transcript of a particularly bad grade in a course they should have done well in. This type of experience is common to many college students, and I’d like to contribute to a project that aims to help this group of students, so that they don’t have to go through the same stress and worry.

In high school I often talked with my peers about concepts we were learning about in class, and although the talks were fun and interesting, it would be frustrating at times when it felt like these students had a selfish motivation of only gaining a surface-level understanding of the topics in order to pass the next test, only to forget the material later and then go through the same process a few weeks later at the next exam – it didn’t feel like they were genuinely interested in the material. Being a procrastinator myself, it’s difficult not to empathize with or even excuse parts of this behavior, especially when they show that they’re hardworking and capable in other areas – working on your weak spots is difficult to manage when you’re not sure what your weak spots are, or how to improve at them. But it was rewarding – and even inspiring – to see some students go from disinterested and underperforming in a class to engaged and acing assignments after watching 3blue1brown videos on youtube and reading some of paul’s online notes: I have personally seen the change that can happen when a student finds resources that genuinely interest them and how when they start to improve, the snowball effect it can have on them. That’s why I’m personally staked in this project – because I’d like to be able to provide that same experience to any student that might be under similar circumstances. From my experience, students tend to

be the most engaged and productive when they're not under the selfish mindset of completing an objective by some timeframe, but of interest in the subject itself.

Because of the nature of the CS industry, it can be difficult to instill the skills that colleges set out to teach to students without a clear goal in mind (background theory on the subject, general problem solving, the go-getting skills needed to pick up new technologies on the spot), and the added responsibility that falls on students in college, it can turn out that a lot of the work falls on the student to self-learn certain topics and gain experience first-hand. But it can be difficult for these learners to find out where they need improvement and how to gain experience in these areas to overcome these challenges. There's an endless supply of information, resources, and topics accessible to students on just about any topic that they could be interested in, and providing them with an interface that helps tackle that obstacle. Students at Risk helps bridge that gap by determining at risk students, and providing them with resources that can help them – and hopefully get them more engaged in a topic and bring them out of being at risk. Students at Risk can do this by providing a quantifiable metric for determining performance as well as providing actionable feedback to each student, personalized with the use of (ai/ml) to help them specifically in what they need more work on.

Ideas

I think it's important to break down the problem into two subtasks: first determining which students are at risk in a topic, and then to provide that student with the resources that they need in order to improve. Canvas provides an API – Canvas LMS – which can be used to gain student data (performance on tests, quizzes, homework assignments, etc.). After ensuring that this data is FERPA compliant, we can use it to determine students at risk. The data will then be extracted for features that will be used to tackle the subtasks. What sort of features we could try to extract from the students and how could it affect the model? Should they be student-oriented? Does the history of a student accurately predict their future performance? I.e. is a student that was previously at risk have a higher chance of being at risk for a future course? Especially if they cover similar topics? Should it be course-oriented? Would a difficult course be prone to making even long-standing non-risk students at risk? Organic chem for instance, has taken from many people 20+hour weeks of study to only barely pass, at UCF and several other universities. Or should it be more granular, at a lower-level scope, where each assignment is tested against, predicting the students performance at that assignment? Particularly difficult tasks might be

outliers and skew the data when all students perform badly or a student that was once not at-risk suddenly becomes at risk. These assignments will come up and we'll need a way to work with it. One metric for predicting a student's performance could be logistic regression, as it models the probability that a discrete outcome happens given a set of input variables. This could make it useful for this task, but it's also not very accurate at non-linear or more complex relationships, which is likely the case with this project. We could also try Random Forests, which are better at modeling complex relationships but tend to be more computationally expensive (as well as time). The binary of "at risk" vs. "not at risk" might make Support Vector Machines a viable option, as they tend to work well in higher dimensions – which would be useful depending on the scope that we pick in the project, this option would be versatile. But the drawback is that they don't provide probabilities – they're a discrete classifier, but that might make testing easier when it's easier to correct a model when they give an incorrect output than to grade it on a continuous scale. A standard CNN might be useful in determining students at risk as they're very generalized and can pick up on non-linear relationships, but they're also expensive in terms in computation and time. Most of these approaches would be too expensive to run locally on GPUs, so we'd probably have to look to training models on the cloud – Google Cloud ML Engine provides a platform to train models at scale, but can be costly depending on the compute hours and data storage needed for these models. AWS also provides SageMaker, which is a similar cloud training platform that also gets expensive depending on the use, but seems to have a more in-depth console with a lot of fine-tuning options made specifically for data scientists. It will take more looking into to determine which option provides the most efficient use of resources for training models.

Jordan Morillo [Backend]:

My Motivations:

This project resonates with me because, as a student, I can say most of the excessive stress during your education is tests and quizzes (and then probably projects as a close second). College is filled with grading scales that will put you at a massive disadvantage if you fail even one test. However, some professors are very secretive about what's on their test. Is it going to be like homework? Other quizzes? If the class is standardized, it might not be even like the lectures. Sometimes the only solution is to just obtain as much knowledge on the subject as possible, but you might not know where to start. As the saying goes, you do not know what you do not know.

Let's say you failed the test, and you want to improve. How? Some professors don't show the correct answers after the test, and you'll be left wondering forever what you didn't know, especially if you were confident on that assignment. Speaking of confidence, sometimes you see a test coming up, and you think to yourself "I can handle that one, I'll spend more time on other subjects that need more of my focus." However, on the test day, you realize you truly know little on the subject, and should've spent more time on it. Haven't you ever felt that a second perspective could've solved all these issues, but you may not have time to get a tutor or be available during professor's office hours.

These are all problems I and fellow students have had, and I believe is the source of most of students' stress. I believe this project can solve all those problems. Don't know where to start? The generative AI in this project can recommend good starting points. You're super confident in the next test so you won't study? This program will be able to warn you that you are predicted to fail, so you spend more time than you usually would've. Failed a test, it's been graded, but the instructor doesn't post any sort of explanation? The application may not be able to give the exact answer, but it will send remedial content, and while looking over you may be able to realize "Yeah, I'm weak on this subject" or "Oh, I didn't know that's how I was supposed to handle that question" and adjust accordingly. The best part is that the project will do all this for free, and you can access it whenever you want, alleviating the burden of office hours and tutors. The idea of creating a project that other students can use, and that the freshman year version of me would have thanked me for, in the same way other students thank programs like Quizlet for passing, is super rewarding and motivating.

My ideas:

Earlier, we discussed having the need for a database to hold the student's data, instructors' data, and the assignment data. I have been tasked with researching and picking out the best database to complete this task. On the relational database side, during our research, we think using MySQL for its performance and scalability would fit the amount data I expect that this program will be accepting. It also has the benefit of being open source, so our internal testing wouldn't affect our budget negatively. On the non-relational side, I have some experience with MongoDB, and I believe it would also fit our project. The flexibility and usability of Mongo make it a great contender, and it is also open source, keeping our internal costs low. However, it is known to not be efficient when given an excessive amount of data, which could create issues

in performance down the line for our sponsor, on the other hand, while MYSQL is more efficient on paper, creating a schema for it that saves an unknown number of different topics could be complex. For right now, we will use MongoDB, but I will continue researching for an efficient schema for MYSQL, as I believe it will save us on performance in the long-term.

The potential need is to find a webhosting service. There are many great web hosting platforms. I've looked into a2hosting, a webhosting platform that has excellent reviews online as a cost efficient, reliable, great scalability and having support for LMS' (Learning Management System). Considering the current plan for this project is to be an application that is integrated into Canvas, an LMS, I believe this would be our best option. It also has great scalability, to have backups, I have also looked at alternatives such as Amazon Web Services (AWS), which also has great review for being reliable, and is widely said to have the most available services of any webhosting platform. This comes with the downside of being egregiously expensive. Another alternative is Azure, being a middle ground in price between a2hosting, and AWS, more powerful than a2hosting, however having less services than AWS. This makes Azure a great second choice, and AWS as a luxury choice.

As the current plan is to save student data into our database, it also means I need to implement security measures to be FERPA (The Family Education Rights and Privacy Act) compliant. I understand that educational records hold PII (personally identifiable information), and that student privacy is federally protected under FERPA. To maintain security, I will focus on following the CIA triad model, focusing on our saved data being confidential, maintaining our data's integrity, and allowing it to be available when correct credentials are inputted to access it. I will be practicing security measures such as having authorized roles within the database, which will help availability and confidentiality, as it will only allow people with appropriate credentials to see sensitive data, such as the instructor. I plan on hashing these credentials within our database, to add an extra layer of protection. I will also investigate implementing intrusion detection, through the act of logs and other security measures within our system. This will help the project maintain data integrity, alerting us that there may be a potential breach of data protection. Additionally, this allows us to notify individuals that their data may have been compromised.

After I implement all those security measures that are deemed necessary, I will start looking for additional implementation that will continue to bolster security. I will be looking into

encryption of student information, instead of just the passwords, which mitigate data breaches even further. Another option is implementing mobile authentication, having instructors answer a notification on their phone and a password will mitigate intrusions within the database. Additionally, I will safely discard the training data for our machine learning algorithm after training.

Ramir (Backend): It was amazing seeing so many great projects being pitched this semester. At first, I struggled to narrow down my choices, until we got to “Student at Risk Predictor”, and I was like “that’s the one, that’s my project.” I rated it to be my top 1 project, and I got matched with it, how convenient! At first while choosing this project I did not actually have a specific reason for choosing it.

Doing the personal motivation of this project got me into reflecting, so I got to retrospect and asked myself what is the Why for rating this project so high. This took me way back maybe to around middle school, all the way through high school and even some college classes. The common themes are test anxiety, the desire to perform better in some exams and so on. As a student, it’s easy to get stressed by some topics in a class, therefore by being stressed it can be hectic to devise a great study plan and be successful in said class, or at least this has happened to me. Now, I know some would say office hours, TAs and whatnot exist to help through the class, but one counterpoint I have is that it would be in some instances hard for instructors or TAs to pinpoint which aspect or topic of the class one should focus to excel better, as they already have so much on their plates. This project is about solving this issue, as the Minimum Viable Product (MVP) as of right now is to provide personalized feedback in a way. Back in the days, I would have thought such a thing was magic and now I get to be part of its very creation, which is very exciting. We are about to revolutionize things here. I must also say that I will be very sad that I am not going to get to use this in my college journey, it would have made me a much better student.

One other reason that motivates me for this project is the different components that it involves that will make it work. I always wanted my Senior Design project to involve some sort of machine learning. The nicer thing about this is that there is also a need for a complex backend, which is one area of interest of mine, so those things picked my interest and made me want to work on this. I am also happy about the breakdown of the roles, everyone on the team has gotten the role they wanted to work on, and we are also attempting to have a system of rescue or a “Bus

Insurance” in case a part of development is suffering for one reason or another. Those little things motivate me way more to have been partnered with thoughtful teammates who are excited as I am to work on this.

We all look forward to having this project done and working by the end of 2024. We will be doing our best to get this as operational as possible while we also polish the User Interface (UI) to provide the best User Experience (UX) possible. We also look forward for this to be used by actual students soon so it can start helping them study better and possibly boost their performance in each class.

API (Application Programming Interface) and Endpoints

Now that we have gone over some of the goals we hope to achieve with this project, now is the time to discuss how we plan on making all these possible. First off, we can't make this possible without the proper tools. We have been exploring some technologies to develop the API, or the system that will establish the communication between the users and our database, we have not fully decided yet, as we are still exploring and researching which one would be the most appropriate for the scope of our project. We also need to be careful and thorough with this step, as our app involves Canvas Integration and communication with some machine learning models, we need to make sure that we choose a suite of technologies that will make the process seamless. So far, we are thinking about building RESTful API and the different technologies we are considering so far are Node.js with Express, potentially with TypeScript to ensure type safety or Python with either Flask or Django. So far, we are leaning more towards Python with one of its frameworks, as we are trying to limit the different languages we will be using.

After having properly gauged the scope of this project, we will be able to choose which Python framework to work with, as Flask is good for smaller scale applications, whereas Django is oriented towards bigger scope ones. That is of course if we choose not to pursue with Node.js, those are still up to discussions within the team and our sponsor.

Now, what we are trying to accomplish here is to come up with some RESTful endpoints that will facilitate the communication between our machine learning model(s) and what we show to the user(s). We may also rely on endpoints to know who's a professor and who's a student. One of the first endpoints we will probably think about implementing will be some sort of login endpoint. Since this will be integrated within Canvas, we were thinking about not having the

users to explicitly create an account, but having that created for each student, provided that the instructor of the course will be using our software. We will then need to rely on some endpoint to retrieve the Canvas user login and connect to our app using this login via maybe OAuth. Some issues may arise from that, but we will discuss those in another section. Some endpoints will also be created to ensure students' information retrieval, those may have some issues related to them, but we'll discuss them in another section.

Lastly, to test those endpoints and ensure that they work we will be using Postman, we will also be documenting each endpoint with something like SwaggerHub to ensure that everyone on the team is able to use them without any issues. Next thing that will be up for discussion on this will be the hosting platform for the API.

Potential Issues

Now let's talk about some issues that we may encounter. Over the last section we quickly mentioned some issues, such as using the Canvas login to connect to the app we are developing, and we also mentioned some issues that may arise from the retrieval of students' data. Students' data may include names of students, their grades, and all other private information. We must first research and understand if the reusability of Canvas login is not covered under FERPA and if it is how we can proceed to either use or if we'd have to create another account creation mechanism. The benefits of this proposal would be that we wouldn't need to explicitly have those schemas for a database, as they would already be accessible through the Canvas API, it would facilitate the students to access their accounts to quickly check their standing in the class (whether at risk or not) or to review any suggestions the algorithm may have provided without even having to worry about an email and a long password.

To get this done, some of my initial thoughts here were that we could have some "login" button that would use the student's Canvas login to proceed to our app, but after some thorough research this could lead to multiple security or even legal and compliance issues. A better approach would be to use Single Sign On (SSO) integration with Canvas and in that way, we handle the login with the login session of Canvas. Now this will come with some other challenges, such as managing some tokens, and some security issues, but if we choose to go with this route, we will be thoroughly testing it. We will also need to provide good documentation for the students, so they understand how this works and how their data is being used.

Another issue mentioned would be the retrieval of students' information. This is also another thing that we must be extra careful with on how we proceed because we must ensure that students won't be able to have access to other students' information whether that is through the SSO or when the data is being retrieved. One way that could possibly be a fix for this would be to maybe map each student to a specific key, or maybe work with some token, JWT maybe? Ideally, we would want to work with the SSO we would put in place to somewhat have a key to map with a distinct student to retrieve their own information. Another issue would be the possibility of different permission levels having different levels of access. Let's say professors have a level of access that would probably be the highest one, if the course will have Teaching Assistants (TAs), they have a lower access than professors, but higher than students. Those levels could probably have access to all the students. Then the lower level would be the students where we make sure they can also see their own records and their own report. This would be an area where we must be extra careful on how we come up with it. As of right now, we do not have a set way for it, but it is a discussion point.

Team Contract:

General Expectations for every member:

- 1) Presence:** Each member of the team must make a considerable effort to attend meetings and lectures as much as possible. They must also be accessible via Discord, our chosen communication platform, and immersed in team practices as they are established throughout the progress.
- 2) Independence:** Each member is expected to work on independent tasks without compromising academic integrity, and their independent assignments completed to the best of their ability without over-relying on other team members to complete project tasks and assignments.
- 3) Collaboration:** Each member is expected to collaborate on role overlap and rely on each other's skills and experience where appropriate. For example, front-end and back-end are anticipated to collaborate for integrating and calling Canvas APIs throughout the project. There will be no "sectioning off" responsibilities over semantics of roles.

4) Timeliness: Each member is expected to respond to messages in Discord and attend meetings in a timely manner. More details regarding attendance and sick day policies are listed below.

5) Communication: Discord is the team's main mode of communication, including the sponsor. Members will also exchange phone numbers to increase communication in emergency events and lack of correspondence on Discord. Members are also expected to give notice for events such as trips and work obligations.

Role-specific Expectations:

Project Manager: The project manager is responsible for managing and facilitating Agile (Scrum) practices including Sprint planning, standup, and backlog grooming. He is also responsible for facilitating communication among team members themselves, and the sponsor, as well as resolving or mitigating time or interpersonal conflicts. Other responsibilities include scheduling meetings, organizing group work on assignments, and establishing deadlines and deliverables for each team member to complete.

Frontend: Our frontend developer is responsible for creating and deploying an interactable user interface. Frontend will also be expected to design prototypes and incorporate user/team feedback into UI development. An integral component of this project is illustrating the risk of a failing student to an instructor on Canvas. Options on color coding, using a dynamic meter, and similar portrayals will be explored in the first semester.

Backend: Our backend team consists of two developers to integrate APIs (application programming interfaces) for Canvas and facilitate the flow of data between different facets of the application. They will interface with other technical roles frequently to help query APIs and manage data across the different application layers.

AI/ML: This specialist is responsible for training and testing a machine learning model (tentatively via regression) to predict whether a student is at risk. AI/ML development will also work closely with the sponsor regarding mining data for use in training the model, adhering to ethical guidelines surrounding the anonymity of "at-risk" data. Finally, the specialist will expose any custom models, so they are usable from either server or client side and by extension accessible to the end user.

Meeting Policy: Sponsor meetings are typically held on Fridays at 2PM ET, on Discord calls. Any per-week changes should be communicated 48 hours prior to this meeting. Due to the team's widely varying availability, most team calls without sponsor are slated for Sunday mornings but are often scheduled on a circumstantial basis due to drastically different schedules. A day's notice is typically expected of these meetings. Meeting agendas are drafted and posted to Discord by PM, after collecting specific questions from members in DMs. Notes and Lessons Learned are documented by PM and posted in the #standup thread in Discord. The PM will take attendance at the end of mandatory meetings using Suivix, a Discord bot that can generate an attendance sheet for each meeting.

"Sick Day" Policy: In case of last-minute attendance changes due to emergency (family, illness, work/income related, please contact another team member via phone call. As per class policy, if a member cannot make an in-person check-in with instructors, the team will try their best to accommodate within reason. Any expected absences should be communicated as soon as a member is aware of them.

Accountability Policy: Members should communicate if anyone is offloading work, they are responsible, and accountability will be supported through peer reviews. There will be no falsification in peer reviews, and each member will be honest about the achievements of their own and others. Status check-ins are situational depending on the task at hand, and members are expected to provide them consistently for each assignment as agreed upon by the team.

Name: Justin Gamboa Signature:  Date: 2/9/2024

Name: Jordan Morillo Signature:  Date: 2/9/2024



Name: Jason Saini

Signature: _____

Date: 2/9/2024



Name: Ramir Dalencour Signature: _____

Date: 2/9/2024



Name: David Santamaria Signature: _____

Date: 2/9/2024

Name: _____ Signature: _____ Date: _____

Design Proposal

Executive Summary:

Objective: The objective of this project is to develop a tool integrated in Canvas that will assist instructors in identifying and supporting students “at risk” of not passing their courses. We will leverage both newer and more traditional AI methods: custom generative GPT models and predictive algorithms such as regression.

Methodologies: This project utilizes a concise tech stack to integrate with Canvas. [TODO: Narrow down tech stack]. Our approach involves the use of regression to predict at-risk status of a student based on training data from previous non-passing students. This will be provided by our sponsor who has experience data mining performance data from students on Canvas. As per ethical guidelines, training data will be anonymized. This project will also leverage Agile Scrum to develop iteratively and assign tasks for each team member (detailed below). Everyone is added to the JIRA board and once we narrow down an MVP, we can start populating it with

tasks. We already have a working concept for a custom GPT that generates tailored YouTube content in a programmatic JSON format.

Result: The intended outcome of this project is a reliable and scalable Canvas integration that can predict at-risk academic performance and provide support in the form of instructional content recommended by generative AI, in the form of YouTube videos. Instructors can step in and help students in real time with this integration, adding a preventative measure to failing a class.

Implications: The potential implications of this project would be quite significant to the modern academic landscape. This project would enable instructors to intervene early and increase the rate of passing students across the board, leveraging targeted content to forward improved academic success. This tool will evaluate students' weaknesses individually and enable personalized support.

Future Directions: While the above details provide aspects of the minimal viable product (MVP), we have also established some reach goals with our sponsor. Reach goals include feedback integrations to improve GPT's content recommendations and letting the instructor select specific videos from the automatic content generation. It would also be wise to quantify and analyze the efficacy of this project before potential scaling.

Frontend

User Interface Design & Interaction:

The design for the tool is meant to be a low-profile tool that naturally enters the viewport of the user. The goal is to have an interface that complements the rest of the Canvas UI in addition to having it's own distinct features to be able to stand out. Interactions will be minimal at first but further functionality will be available.

Visualization Strategies for At-Risk Metrics:

The methods of describing how at-risk a student could be broken down into a couple methods:

- Percentage out of 100%
- Color slider caution symbols

- Green being the safe zone
- Yellow meaning there's enough cause for concern
- Red meaning they are failing currently and need assistance

Discussion on Canvas Integration vs Entire Website

There are tradeoffs to having a fully integrated tool in the Canvas portal vs an external website.

Starting with the largest benefit of having an integrated tool would be a more fluid and seamless experience that does not require an external site and improves the workflow of the user. There are tools that require an external site, for example an FAQ section in an app that takes you to a separate browser that immediately diminishes the user experience which often results in the user scrapping the tool altogether. The goal is to lessen the friction in using a new feature and with the extensive documentation that exists for a goal like this could be useful.

However, as previously mentioned there are tradeoffs like the greater technical requirements needed to integrate an app into an existing application. Creating a website would significantly cut down on development time and could grant more opportunities to fix bugs and deal with any roadblocks we may find. In addition to creating a separate website, there would be full control over the customization of the UI and the site rather than having to adhere to the restriction and limitation of Canvas and its policies. Canvas has restrictions on building apps for the portal which is locked behind two types of accounts: Free-For-Teachers & Administrators.

Administrators have the ability to integrate apps at the account level meaning they don't need to have any courses to add apps whereas the Free-For-Teacher's account allows for app addition at the course level meaning that they can add apps into individual courses. Either approach should not prove difficult, however the only obstacle that could prove troublesome would be how to secure funding for the administrator account.

Backend

System Overview

We have been tasked with the development of the project "Student at Risk." The project is an application that will, via machine learning and generative AI (Artificial Intelligence), do two tasks; primarily inform students that they may be at risk of failing their next quiz or test and

recommend resources that may help them avoid failing. Secondly, it will provide remedial content for previous failed quizzes and tests. The current plan is to have this application be directly integrated into the Canvas, a pre-existing application that allows for the assignments, quizzes, and tests to be virtually distributed to students, all under the control of the instructor.

To complete this task in the project, we will use the Canvas API, whose documentation is publicly available on the Canvas website. The API has endpoints that will allow us, with correct verification, to obtain vital information about any individual course, such as its assignments, the students enrolled within the course, and the instructor teaching the course. We will collect data using these end points, focusing on assignment's topics, student performance on assignments, and student's current grade. We will create a parser that will take the given data and make it into an easily readable and savable data. We will also need to save the instructor's verification key, which allows me to access their course data through the API.

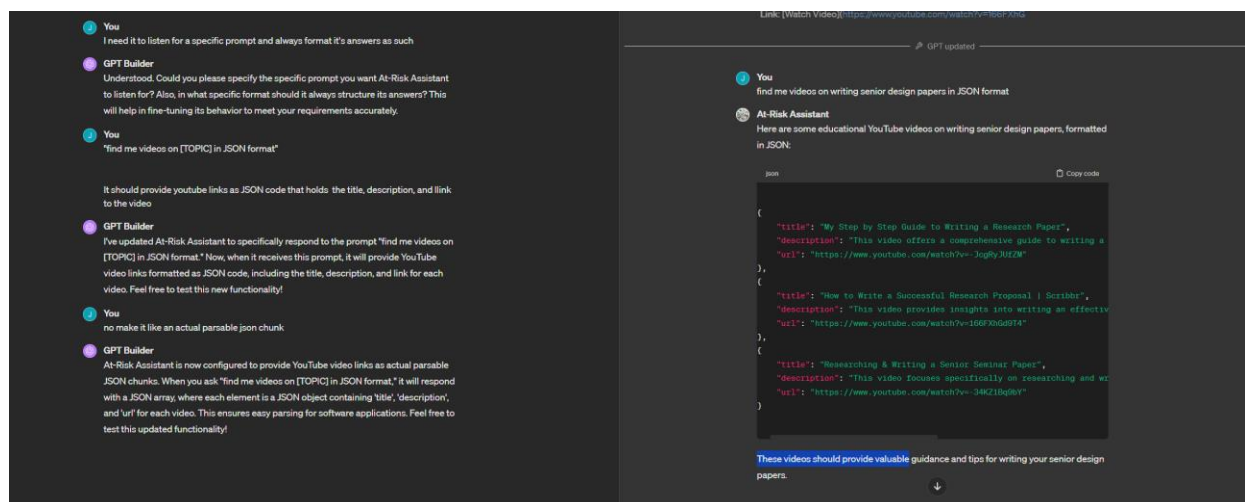
Once this information is obtained, we will then save this data into a database, filled with tables, which will be populated with the student's data. This data will be fed into the machine learning algorithm, and by using linear regression, will predict the student's future performance. This prediction will be compared to real, anonymized, grading data provided by our sponsor, to fine-tune the machine learning algorithm. When a student is predicted to "fail" (Below 70% correct rate) on the upcoming test or quiz, they will be deemed a "student at risk." This will cause the second part of the primary function to start, which is recommending material that will hopefully avert the predicted outcome.

We will complete the task of recommending material by taking the predicted data, and the topics of exams that were "failed" and send it to our chosen generative AI, which will mostly likely be ChatGPT. The generative AI, via its API, will be given the failed topics and be prompted to, in a friendly manner, recommend YouTube videos and clips that may help the student in the subject. This generated recommendation will be then sent to the student and the instructor. The instructor will be given a choice to override the generated recommendations with his own curated responses. The instructor can view whether the student clicks the link or not. As a secondary function, the program will also provide remedial content for previously failed quizzes. This will be created with a similar process that is used for recommended videos to avert future failing. The application will have a function that if a failing score is inputted into the database, it will send that topic and student information to the generative AI. The AI will do a similar process of

generating a friendly recommendation on the failed material, and the instructor will have the same choice on whether to override the recommendation.

GPT YouTube Recommendations:

Alongside our traditional regression approach for predicting students' risk levels for failing a class, we decided to provide curated content on target subject areas based on assessment performance. We can build a custom GPT through the OpenAI web interface, as shown below. The link in my motivations section requires the purchase of GPT Plus Subscription.



One important consideration of utilizing ChatGPT from an API perspective is the cost. It typically costs a cent for every 1k tokens (words). The words are not necessarily a one-to-one ratio. Our current prompt, "Find me videos on {TOPIC} in JSON format" is 9 words in the context of GPT tokens. If we were to test our custom GPT on an API this is a consideration for every time we query it.

An important consideration is the handoff of this project. Hopefully for future use by institutions, this cost will be ongoing and affected by the GPT marketplace. There will be a need for a business analysis to assess the dollar tradeoff of providing curated content. This solution's efficacy will also need to be measured from a student and instructor perspective to determine if it is worth implementing on a large scale. For example, scaling this for UCF would be expensive depending on how many at-risk students there are.

Another development we must consider is how we are going to query this API for multiple topics. We need to understand how questions are organized on assessments and find a general-purpose way of querying this GPT model for multiple topics, as well as making sure we can pull that information from Canvas reliably and consistently.