Theme	Code	Quotations GF1			Quotations GF2		
Adequacy of the branching model	Navigation between the model branches is straightforward.	S1-1.b-P4: "It's calm Because... It's... It's... Experiment And the issue tag And its title So... It's very... Calm." S1-1.b-P1: " (Moderator): Closed And so, thinking about this structure, would it be easy to navigate? Is it possible to understand what each one is? (P1): Yes.			(...) "the last question here, from this first session, is about whether you were able to understand the instructions about the structure itself, of the model branching, of the branches and on the criteria that were proposed there, gradually There. To be applied between the transition of the environments, and between the branches. If you, when you read it, was it clear to you?" S1-4-P2: "For me, yes"; S1-4-P3: "Yes"	S1-1.a-P2: "So the branching was pretty clear from the beginning"
	Branching model is suitable for versioning AI models and datasets	S1-1.a-P4: "So, I didn't see... We had this... When this model was made available, we had several discussions about how it would work, right? When we started experimenting and seeing how it would fit into our development, I, personally, realized that it doesn't make sense. for the experimentation part. I felt that... Because... When we do... Experimentation is not related to what we will make available to the team. We... For example, I work on experimentation code, training and everything. But that is not what will be made available in the stage, for example. What will be on the stage would be the resulting model." S1-1.a-P4: "Yes, but I think the team, the other teams that would consume, they have no interest in the training code. This training code my experiment, you understand" S1-1.a-P4: "For example, other teams will have access to things that they make no difference." S1-1.a-P4:" (P1): But they want to track your training, right? (P4): Yes, yes. So, I think this should be separated from what we are going to serve. For example, the service that we are going to provide, this may be what will be consumed, along with the model that will be generated by the experiment. And this has... I think this has to be separated from that we... From the garbage code that we're doing. Like, from... Garbage like that, like, is disposable for other teams. They don't need to have that access." S1-1.a-P4:" (P4): Because, like, like [P2]. [P2] will do it in the same place as us, okay? The AI Experiments that we have today. What is our experiment today. [P2], are we going to do the service there? What will be consumed? (P1): No, it's another repository. (P4): That's what I'm saying. Like, they won't consume from the experiments."	S2-3-P4: "but for the context of model and dataset versioning, doing this triangulation, I think it works."		S1-3-P2: "And then I feel that this set, if we were versioning the model, it would already be, it would already be very good, you know? It's good enough to version the model.", S1-3-P2: "Like, in the model I see this flow following better like, oh here I have an experimental model, here I have a model like the development version, here I have a model that is in stage, here I have my main like this, it was. But without versioning the model I I have a little difficulty, I'll be honest.", S1-3-P2: "Yeah, I think if it were accompanied by a versioning of the model, following the same standards, maybe it would be... it would be easier. Or if the model was versioned in this standard"	
	Branching model provides a suitable framework for AI model evolution for staging and development environments	S1-1.c-P1: "'Well, my opinion is very suspicious because I already work with that on a daily basis, at work. So I had to implement that because of all the problems I had. In fact, work Who does a merge there in develop, in stage, I have code review, right? I have other people looking at it. And, for example, there in main there is only one person who throws things into production. Got it? So, approved, approved, but it is not the developer who puts it into production. It is someone else. So, from a business point of view, it makes total sense. From an AI point of view, I have no opinion." S1-1.c-P4: "If the service, and if this is to serve as a model, I think it makes sense. If for a model, I think it makes sense, for a service, I don't think it does."					
Lack of motivation to use the branching model	Branching model restricts experimentation		S2-3-P1: "Yeah... and then I ask you, do we have the option of having this flexibility in the experiments part? So, no. You want us to use the model."	S1-2-P1: " I'm seeing that it's very similar to things I already do. I understand everything involved in the story, but reading the article once, trying it out, I will do very little. I don't know what balance we will find in the project, to meet all these other sides, because in the end we You need to create a product, but everything you said is good practice, it has to be done, it would be great if you could have done it."			
	Branching model not used for artifact versioning						S1-1.a-P3: "But, at least for me, in that first sprint I didn't... I I was just creating artifacts, I didn't have actual code, so I wasn't using the repositories." S1-1.a-P3: "I generated a PDF document that has all the results, all the prompts, all the comments, but it's really just in one folder that we have in Drive, it's not in Git yet."
	Branching model was not used due to the possibility of adapting the model		S2-1-P4: "So I didn't worry about following the model, because I thought it would be feasible to change."				
	Versioning experiments from the branching model must be performed due to third party wishes	S1-1.a-P1: "But that's what they want. They want you to open a branch when you start doing a new experiment, you know? Name it experiment, and then you'll get to the end of the sprint and you'll have 5 experiments, 10 experiments, the one that had the most results, you will merge to develop." S1-1.a-P1: "But this experiment... You will version your code. Your setup. Yes. The result of the model. It's not supposed to be versioned. But what they want is the tracking of their experiments. Of their code changing. The parameters changing." S1-1.a-P1: "It's new for us. But I think that's what they want. You'll generate the experiments. When we get here and present the results, some of these branches will be chosen to go into development, production, etc. Then that one will be merged. And the others will be closed." S1-1.a-P1: "But I think we would open new branches at the same time Classifier repository and over time the code will be replaced, changed, you know? This will be gradual. But what they want, and we are part of the research project that is bigger than the AI people, is how an AI team adapted to use a CI/CD flow, git control, versioning, within our experiments. So that's what they want." S1-1.a-P1: "But they want to track your training, right?" S1-1.a-P1: "Okay, I get it. That's the point. If you spent three months working on a series of experiments And you're going to hand it over to your boss And none of it worked out He'll probably ask you where that these N experiments that you did are versioned. Got it? I think that's the point. Got it? I think that's the point. So, if it's git flow, if it's this, from here, it doesn't matter, but it's... What's a given for us is... We version our experiments, and know that a series of things will be closed, will remain in the past. Got it? Now... If it's git flow, if that's what they're proposing, I don't know. Because sometimes you're proposing this here based on an article you want to publish. Some reference. Got it?"				

Theme	Code	Quotations GF1			Quotations GF2		
Need to adapt traditional branching models to experimental environment	Branching model is very similar to Git Flow	S1-1.a-P1: "Okay. Why don't we use Gitflow?" S1-1.a-P1: "I don't know, I don't see any difference. You have a master that is the main. You have a hotfix that comes and destroys everything." S1-1.a-P1: " (P1): So, I'm going to create... I need to work on a model to increase a metric. Then I open a branch. I work, I do experiments there. It generated one or more branches, there. It doesn't matter. Then I get the result. We make the decision. Are you going to merge or not? You're going to merge. You're going to develop as if it were the user's issue. Aren't you going to merge? We'll close it. And it's saved there in the history. (P4): I think there's also a lot of semantics involved. So, when we're doing experiments, we don't work on a feature for the product. (P1): Of course I work. If you are doing an experiment based on some expectation. (P4): Yes, but... Yeah, that's what I'm talking about. Like, the feature will be an addition to the product. Not necessarily an experiment. (P1): Oh, but that's okay, but that's the thing. It's not a problem. (P4): But that's not GitFlow. Got it? That's like thinking in semantics. It can have more or less structure there... If we rename feature to experiment it becomes the same thing. "	S2-3-P1: "From what little I know, as I said, it looks a lot like Git Flow. It's right up my alley."				
	Despite the similarity, the original Git Flow may be unsuitable for the experimental nature of an AI project.	S1-1.a-P4: "Yes. But the issue is one thing. Gitflow is more based on features. It's about functionalities, right? We don't make functionalities." S1-1.a-P4: "No, I think it doesn't make sense in GitFlow. What GitFlow has, if you use it, is... A lot of branches that are discarded." S1-1.a-P4: " (P1): So, I'm going to create... I need to work on a model to increase a metric. Then I open a branch. I work, I do experiments there. It generated a or more branches, there. It doesn't matter. Then I get the result. We make the decision. Are you going to merge or not? You're going to merge. Go to develop as if it were the user's question. Are you not going to merge? We close it. And it's saved there in the history. (P4): I think there's also a lot of semantics involved. So, when we're doing experiments, we don't work on a feature for the product. (P1): Of course I work. If you are doing an experiment based on some expectation. (P4): Yes, but... Yeah, that's what I'm talking about. Like, the feature will be an addition to the product. Not necessarily an experiment. (P1): Oh, but that's okay, but that's the thing. It's not a problem. (P4): But that's not GitFlow. Got it? That's like thinking in semantics. It can have more or less structure there... If we rename feature for experiment it becomes the same thing.					
	Hotfix is unnecessary for the context of experiment development S1-1.a-P4: "Hotfix is unnecessary"						
Flexibility and adaptation of the branching model	Experiments should be versioned in repositories with a less rigorous branching model	S1-1.a-P4: "For models and... And... Datasets that... That come from experiments. I think that works quite well. This... This... This model that you made. And for... For the experiment... I think it should be something closer to a... A... Trunk Based. Because it's... These are quick things that we... Sometimes we run in... In two days there... Oh, it didn't work? Okay. Let's... Let's leave it registered. But it won't... It doesn't need to have such a... big progression on top of it. But on top of... Of models and... Datasets and... And... Services that need to be tested normally. Think you need it." S4-1-P4: "and for experiments with the experiment codes, the resulting artifact would be reports, it's more like benchmark comparisons, use something simpler, it doesn't need to be trunk based, but maybe it isn't so stuck to the guidelines like that."				S2-1-P2: " And I think that... I think it was a bit like this idea that experimentation needs to be freer, I think. Sometimes without following so many criteria. Like, I do something that isn't working out yet, but sometimes I go up because there's a section, or some idea that someone else wants to take it, like that. Anyway, we followed this instability a lot because it was very difficult to define when something was good, or when it made sense to us, because we didn't know what made sense." S1-3-P2: "I think it's really more like where we are today. The experiment is isolated and the code I'm uploading is already a code more oriented, therefore, towards something."; S2-1-P2: "And then we want... that the creation of this service is a service that serves... and communicates with other teams, and with... with other parts of the code. It should be more modular, as I can change things here, in the experimentation, for the craziest scenarios, and still be able to serve this on the other side. You know? And I think... I think it was a bit in this idea that, experimentation, it needs to be freer so, I think"	S1-1.a-P2: "And then, our experiment repository kind of loses that depth, as if it only had experiment and main, let's say. S1-1.c-P2: "...because it's as if we had also placed another repository there, so it's as if it were divided in half there, there was kind of like ours, our other repository starts from develop there, kind of taking some things from what's coming from the experiment, so I think for us, at least at the moment, the ideal would be like it had a little bit more down there, like, in experimentation." S1-1.c-P2: "...and then I think that in that scenario, we need to adapt a little. Because maybe it doesn't meet everyone's needs" S1-1.c-P3: "that we made these small changes to the two repositories to meet the... The moment of the project."
	Adapting the branching model to meet the needs of the team throughout the project		S2-3-P1: "Yes... probably at some point we will come to you and specify a demand. You will have to serve us in the same way we are serving you, it is part of the project."			S1-2-P2: "Sometimes in practice, when we actually do something, we find something that we see is a problem or that something is missing in practice, but, for now, in theory, from what we read, it makes sense."	
	Need for a hotfix branch in the branching model	S1-1.a-P1: "I don't know, I don't see any difference. You have a master that is the main. You have a hotfix that comes and destroys everything." S1-1.b-P1: "So, I think he's missing a hotfix in this story." S1-1.b-P1: " (Moderator): Do you think a hotfix would be needed? (P1): Yes Yes Or separate, which is what Thiago is talking about What is AI, models And AI, service. S4-1-P1: "Technically, in relation to the model, I think we have to consider the Hotfix, as was observed."	S2-3-P1: "I'm just missing the Hotfix within this flow."				
	Separation of experiments and webservices into distinct repositories		S2-1-P4: "I didn't bother following up because I was hoping to have this meeting to discuss this, because I had found that experiments and services had to be separated"; S2-1-P4: "I don't think it makes sense to have an experiment together with the service that will be consumed."; S2-1-P4: "I think it should be separated from the contexts as much as possible."; S2-3-P4: "But for the context that we, if we were to do what we did, I don't think it would make sense. Because we didn't have any service and I think that should be decoupled."; S2-1-P4: "in the experiment repository we created a trunk-based version and only with the experiment PR policy for dev, which would be in this main, right. That's what we followed."				S1-1.a-P2: "conversation that we also had a little bit there in... there in the middle from some sprint we said that keeping all of this with the experiment and the service code was a bit strange for us. That's when we ended up separating the repositories, right?" S1-1.a-P2: "And then we had several of these... Of these experiments that... If we were to actually use them, we would have to start all over again. And then to... To also avoid a bit of... So we kind of start over again... With a similar purpose inside there... of the same repository and it is kind of lost what service is. We preferred to do this, you know? It's like, like... Two things were growing inside the repository which is the experiment and the... and the service itself, and we didn't think it was very cool. The experiment part becomes gigantic, and then it would be like... Create a volume of things that we didn't need to look at when we were doing the job. And we chose to just separate them" S1-1.c-P2: "...because it's as if we had also placed another repository there, so it's as if it were divided in half there, there was kind of like ours, our other repository starts from develop there, kind of taking some things from what's coming from the experiment, so I think for us, at least at the moment, the ideal would be as if there were a little further down, like this, in experimentation." S1-1.c-P3: "that we made these small changes to the two repositories to meet the... The moment of the project."
Lack of references on branching models in AI	Lack of Software Engineering Practices in AI Articles			S1-2-P1: "Because when we get the codes, we have to talk to the professor, okay? He's aware of the concern. When we get the codes of the published articles, man, the guys don't care. Go look, and see if there's a test. So, we even took a I was shocked, like, when we saw the... talking about testing, you talking about Git Flow. I understand. It's the final product that we're involved in. But, when I look at the academy, at the academy side, that's not what people do."			
	Little diversity of GCS models		S2-3-P1: "No, I think it's ok. Again, there aren't many models out there if we look on the internet on a large scale."				
	Need for reference sources to support comparisons with the proposed branching model	S4-2-P1: "I have. I think you could bring us some sources for us to compare, understand?"					

Theme	Code	Quotations GF1			Quotations GF2		
Search for process standardization for evaluation of experiments	Standardizing AI metrics for experiment comparison		S2-2-P4: "We have the same metrics. We defined the metrics. So.... with that we evaluated on Friday and ah, this is the one that will be used. So, but the two experiments were recorded as versioned."; S2-2-P4: "We generated the metrics, the same metrics, the confusion matrices there and so on and with discussion, seeing the balance, we had decided on a method."; S2-2-P4: "Talking to the professor, we discussed that F1 and CURVAL, which were important metrics for us to evaluate. Confusion matrix as well. Having defined this, we were able to make comparisons and define the criteria for what would be good and what would be bad."		S3-1-P2: "I think some things are kind of standard, like that. We used... we used the entire Classification Report, which is Precision, Recall and F1. We look at the confusion matrix, too. As metrics main ones, they are documented." S3-1-P2: "[P4] likes to use One DB. I don't know if you know it, but it's like a... it's like a dashboard, like... a BI that you can monitor the training of things. I ended up not using it, I left it off, but I generated the classification reports and put them there." S3-1-P2 "I feel like it's kind of a natural thing, like that. Like... it's just that I don't I know that too. [P4]. I think he already has a lot of experience, so the natural thing would be if... if we get a classification report, we look at as many things as possible. I think he brought a lot more metrics than I did. He brought Rock Rock All Six Score, he already brought a lot of stuff there. But I think that, at least these traditional metrics like that, are what you expect when you're comparing something, or when you're looking at the result, at least."		
	Generalization of metrics across different types of experiments			S1-2-P4: "You can generalize, take F1, which is what we see most. Then it depends on the data, the context." S1-2-P1: "Yes, there are a lot of things you can discuss there, you know? There will be a project that will show that F1 is the ideal metric." S1-2-P4: "Yes, we can define it for specific tasks, you know? For example, if we think about identifying a variable. That's a fixed task, we'll always have to see the variable. We can define the metrics that we will evaluate for this context. So, we define, we want F1, ROC curve, execution time, model size. We can make this definition and then we always define these criteria for this task. But then these criteria change for another task. the RAG task, for example. So, we define there what is best for RAG? It's HIT, several things. So, for each task we can make this definition and use it as a criterion. So, the criteria change from task to task. We can do this. S1-2-P4: "For us to search for context, for example. There is a metric for how well it is retrieving context." S1-2-P4: "In the end, at some point, if we want to replace a method, we will have to compare it in some way and we can define this how are we going to compare."	S3-2-P2: "I think the ideal, in the future, in relation to AI, would be for us to have a fixed set of tests. Like a set, you know, well defined that I want to test my model on that set. And then it would be kind of a final test for us, wow, how cool. We go through that same set and have these metrics here, for each one."		
	The same dataset was used to evaluate the experiments in comparison				S3-1-P2: "And one thing we did, too, is to start from the same dataset. Do trains, tests, splits with the same seed, you know? To have a fairer comparison."		
	Diversity of metrics used to evaluate experiments, in addition to classic classification metrics		S3-2-P4: "Yes, we considered the training time, we considered the... the inference speed, right? It's the time it takes to execute, to make the prediction, we considered these two things. The size of the..."	S1-2-P1: "Because it's not just one metric, it's a set of metrics. Sometimes, it is the balance of several metrics or even the execution time of a business that we will make a decision whether to move forward or not." S1-2-P4: "Yes, there is a cost. If (P2) worked, even with a few loss good metrics there, it would be the chosen one. Because it is more scalable, it is faster to be in production. Mine, to give you an idea, took a day to train. And it takes a long time for you to use, for you to make inferences." S1-2-P4: "There's no point in us getting a business, a GPT chat that we can't even run to talk, no, it's getting better. There's no way. So cost, execution time is also something we think about, that we consider a choice." S1-2-P4: "It's good to at least document that this is also considered, in addition to the metrics. There's no point in us getting a business, a gpt chat in life that we can't even run to talk, no, it's getting it right. There's no way. So cost, execution time is also something we think about, that we consider a choice." S1-2-P1: "In our case, it's a financial cost, we don't even look at it, because we don't even have the money to dare to... S1-2-P4: "Yes, cost of... model size, perhaps." S1-2-P1: "Computational cost."		S1-2-P2: "One thing that we don't have and that we do a lot is look at the size of the model, training time, I think these are things that count a lot. for us."; "whether S1-2-P2: we like it or not, he goes a little beyond the metrics, you know? The metrics exist there, we compare them, but we have to kind of to analyze the... Like, what do we lose by using one in terms of training cost, inference cost, all of that, you know?"; S1-2-P3: "there was one that took 5 minutes to answer me and he, for example, got more than 90% right, but another one took 30 seconds, he got 70% right, so we have to keep thinking about these things too, not just the rate and accuracy between them."	
	Difficulty in defining a standard set of metrics that can be replicated across different problems		S3-1-P4: "Didn't use the same metrics, because it wasn't the same problem."; S3-1-P4: "Mine and P2's were parallel in the same problem. So much so that we had the same metrics. Hers wasn't anymore. It was... it was another..."; S3-1-P4: "For me and P2, a set made sense, and for her, if we took and used the same things, it wouldn't make sense."	S1-2-P1: "It's very complicated for you to specify metrics" S1-2-P1: "Yes, there are problems that you will look at F1, there are problems that you will look at accuracy, there are problems that you will look at loss. There are things that the teacher comes up with at the last minute." S1-2-P1: "There are things that, viably, you come, present and are better than the other. Because it is not just one metric, it is a set of metrics. Sometimes, it is the balance of several metrics or even the execution time of a business that we will make a decision whether to move forward or not." S1-2-P1: "Yes, there are a lot of things you can discuss there, you know? There will be a project that will show that F1 is the ideal metric." S1-2-P1: "There is one more category for the problem. Because sometimes the same category in different problems, the performance expectation changes." S1-2-P4: "So, for example, we can determine what the problem is in the context and define metrics for that. Then, for this, we will consider such and such. And then, for approval, we only need to meet this..." S1-2-P1: "...but each article comes with its own metrics sometimes." S1-5-P1: "The metrics there, I discarded." S1-5-P4: "It's the metrics I also found that..."	S3-1-P3: "It's just that, in this case... mine, my experiments, what I was doing, my experiments were very different from [P3] and [P4]. So, I wasn't comparing them directly with theirs."		
Lack of defined process to document the experiments	Experiment metrics results were documented in other software	S1-1-a-P4: "This is the way we are tracking today, for example, when I generated the results Then I had a description saying, this model here was generated with this environment Using this data and these parameters Today it is done more by hand."	S3-1-P4: "They are registered in other software that, when we talk about experiment, it automatically records this. So, within this software, I generated the report from these metrics."				
	Experiment metrics results were documented in issues				S3-3-P3: "I think we left it in the issue comments. I think the Most of it is described there. Or at least it has a link to something external. And I think on Discord too. We were always warning, like, giving updates, like that. Hey, I ran such and such. The result was this. Hey, the new result came out here."		
	Experiment metrics results have been communicated informally among the team				S3-3-P2: "I think we left it in the comments of the issues. I think most of it is described there. Or at least it has a link to something external. And I think on Discord too. We were always warning, like, giving updates, like that. Oh, I ran such and such. The result was this. Oh, the new result came out here." S3-3-P3: "Also in our team meetings, on Friday, we showed the metrics."		
	Metrics used to compare experiments were not documented		S3-2-P4: "Yes, yes. We already knew which was bigger, which was smaller, but it's not recorded."		S3-1-P2: "I feel like it's kind of a natural thing, like that. Like... I don't know either. [P4]. I think he already has pretty good experience, so It would be natural if... if we get a classification report, we look at as many things as possible. I think he brought even more metrics than I did. He brought Rock Rock All Six Score, he already brought a lot of things there. But I think that, at least these traditional metrics like that, are what you expect when you're comparing something, or when you're looking at the outcome, at least."		
	Acceptance criteria related to experiment metrics have been recorded in an issue				S3-1-P3: "a satisfactory result is when it hits 90%... Above 90%. So it is documented in the task, which was to accept with above 90%. And in the artifact, which I generated, it also has both all the hit rates, as well as a time comparison between them."		
Previous experience and perception of good practices	Preference for using the model because you have already worked with similar models		S2-3-P1: "Yeah, that's right. I prefer it that way, that's how I already work. normally."			S1-4-P2: "Like, we've been using Git Flow, we've been using semantic commit for a long time. So, for me, it's natural. You know? Like, I don't know what scenarios are like where people don't work with that. And like, for me, it's always been natural to do this. So I have a hard time comparing what it's like not to do this."	
	Branching model seen as good practice		S2-3-P1: "That's it, what can you guarantee me? Process. So, the probability of dirt entering production is greatly reduced with this. Is it a good practice? Obviously, but there is a cost."			S1-4-P3: "but for sure... maybe at the moment I don't feel like it's speeding up, but Of course, in the future, to revisit everything we're doing, it definitely speeds things up, having things well organized. It's easy to look back and know what was happening."	
	Branching model allows process stability			S1-4-P4: "Speeding up, it doesn't speed up. It guarantees the process, it guarantees correctness. Not of agility, of other aspects, of correctness." S1-4-P4: "Yeah. It slows it down, but..." S1-4-P1: "It's the fact of working in a team, we have to follow certain processes."			

Theme	Code	Quotations GF1			Quotations GF2		
Previous experience and perception of good practices	Importance of versioning experiments	S1-1.a-P4: "Yes, I understand that recording experiments is important. But I don't think it should be in the same repository as what will be consumed..." S1-1.a-P1: " (P1): That's why you have to create a branch, the name of the experiment, even if it's experiment 101, experiment 102, experiment 103, experiment ABC and 999. Got it? But it changed one line of code. You have to... you have to be versioning. (P4): Yes, I'm not taking away the obligation of having to version. S1-1.a-P1: "Okay, I get it. That's the point. If you spent three months working on a series of experiments and you're going to hand it in to your boss That none of this worked He'll probably ask you where the versioning of those N experiments you did is. Got it? I think that's the point. Got it? I think that's the point. So, if it's git flow, if it's this, from here, it doesn't matter, but it's... What's a given for us is... We versioning our experiments, and knowing that a series of things will be closed, will be left in the past. Got it? Now... If it's git flow, if that's what they're proposing, I don't know. Because sometimes you're proposing this is based on an article you want to publish. Some reference. Got it?" S4-1-P1:" (P1): Yeah... I think that's it. I think the code needs to be versioned. Whatever the model, it doesn't matter. (Moderator): Yes. (P1): But we have the ability to search for the history of evolution of that class of code, of that experiment is the most important. So, I think that's it. I see a lot of value in what you are proposing, and then try to apply it together with the team. "	S2-2-P4: "So, but both experiments were recorded so versioned."				
Application of the branching model is affected by uncertainties or lack of experience	Branching model was not used due to uncertainties about its application	S1-1.a-P4: "Yes, I understand that recording experiments is important. But I don't think it should be in the same repository as what will be consumed..." S1-1.a-P4: "Because, like, like [P2]. [P2] will do it in the same place we are? The AI Experiments that we have today. Where our experiment is today. [P2], are we going to do the service there? That will be consumed?" S1-1.a-P4: "Yes, but I see this in the future, when it scales and we have more experiments, the experiment will be mixed with an AI service and that will cause confusion." S1-1.a-P4: "I think it... Within a repository, experiment should not coexist with the service that is going to be built. That's my idea. The What I feel is that this model is great for model and data set versioning. Because we have the experiment part there. So, there I have the models that I'm experimenting with. I can save them whenever I generate an experiment, I commit it there. You can even do this automatically. I commit the metrics of this model, of the datasets that I used. Then, great. Oh, this one is a... I generated... I finished this experiment, it generated metrics and such. There is a benchmark. Go to develop, to stay unified, decentralized. So, ah, this model here was the one that was chosen. These models that we have, this was the one chosen for... With the best metrics. So, it goes to stage and then... So, I see it working for model and dataset. But, for example, the service that we will do is pure development. It has no AI peculiarities. So, I see that we can use Gitflow to do the job. Which is a development..." S1-1.c-P4: "If the service, and if this is to serve as a model, I think it makes sense. If for a model, I think it makes sense, for a service, I don't think it does." S4-1-P4: "I just wanted to reinforce that if you took a look, maybe talk to [K2], I know he has experience with IA and since he knows a lot about GCS, this issue of decoupling things, how we could... if it makes sense, at least if my idea makes sense, because I'm convinced that this is going to cause problems if we do it, combine experiments and services."	S2-1-P4: "I thought the model didn't match what we were looking for. doing. We thought the structure should be different."; S2-1-P4: "I had doubts, we didn't make anything available, so I also imagined that, I imagined doing this, suggesting that a model should be different for different contexts"				
	Team has little experience using branching models in collaborative AI projects	S4-2-P4: "The AI projects I participated in, just research, hackathons, so, it wasn't... So, it was... And also no collaborative AI, right, it was just me, so it was trunk based." S4-2-P1: " So, the projects I had worked on with AI up until then, basically my master's degree, was a solo work." S4-2-P4:" (Moderator): Was that you and the intern who did the experiments? The other intern. (P4): It was me and him responsible for the AI part. (Moderator): And did you collaborate between experiments? (P4): No, we were on somewhat different fronts. Like me and [P2], we needed facial recognition. So, we needed to improve the facial recognition method. Each one saw a different branch, so, how to solve the problem. And then, we... There was no collaboration, like, direct. So, each one had their own method on GitHub, you know? "S4-2-P4: "When I interned in AI and there, they used a Gitflow, like this, half... of the lame" S4-2-P4: "But, when I worked with AI, it was... we didn't have so much of a problem with this... of following models because the models we were going to used were very small, so they could be versioned in the code together, as if it were... like a file, compiled. so, it didn't make much difference, so, we just... just... but the The experiments that I and another AI intern did, we left it local, there was no environment for us to experiment. Everything was in the local environment, there on the company's notebook and if we needed to share it, it was on our own drive. There was no policy for that."			S4-2-P2: " (P2) Me. I did branching, I don't know, like... like a normal Git flow. It was feature, open the PR. Like, it wasn't going to be approved. I don't think there were these layers of... of experimentation. I never had that. Development should have been, but we were already in another scenario, too, like that. So I think it was a slightly more traditional branch, like that, you know? Like... as if I were to change something I was already changing it with a greater certainty that it would become a Feature, even... A fix for something. But at this level, no. (Moderator): I see, it was simpler like that. Do you remember how many branches there were? (P2): We had the two main ones. There was a homologation, there was a main. Then everything that was around that, like, we followed features, refactors. And then everything that came up, we already opened PR for homologation. You know? There were no extra layers. S4-2-P2: " (Moderator): I see, and you did the PR, right? Or did you commit and push directly to these environments? (P2): So, usually it was a direct commit, but it's a small team thing too, you know? Like, sometimes it was a team of three, four people. And then you if it requires approval from two, like there is one or two in AI. If it requires approval from two, I will review my own PR. Then I open it just for me to accept. "		
	Experience impacting the view on the branching model					S1-3-P3: "As I said, in the form, also, besides being my first experience with AI, it's also my first experience working in a team, like, that works this way. So... but for me, it's sense, you know? I can look at the document, and understand that it makes sense."	
	Lack of time to use or adapt the model, due to the initial phase of the project	"		S1-3-P1: "...in six months ask that question again." S1-3-P4: "You don't even need six months, I think less is enough..." S1-4-P4: "It wasn't a very exemplary sprint, because we only did experiments for dev. So, there was no stage, and we didn't make anything available. So, ask in the next sprint." S1-4-P1: "We're going to start rolling that out now." S1-5-P1: "We had to develop it, get a little more hands-on to understand how it would fit into the scenario."	S3-2-P2: "But it's difficult for us to define this set. Because we don't even have that many training examples, let alone a round set of tests, you know?" S3-4-P2: "I think that at one point, too, we were presenting our pipeline idea, like that. So it's all very embryonic, but I think the idea, at least, is to release some patches notes, so, when we go... go to present." S3-4-P3: "Until now [P1] also, who is our leader, there has been no specific discussion about this. Because we haven't reached that part yet." S3-4-P2: " I think... I think one thing is that it's easier to talk to people. team leaders. I don't know if it's a discussion that's worth extending to all developers, like that. I think that's it, but it's not something that we have very well defined, either."	S2-2-P2: "I think that, in relation to the rules for transit, we didn't change anything. But it's what I said, it's as if we had stayed at the first two levels, there."; S1-3-P2: "I think it does, I think that maybe there is also a very large abstraction, because we are not versioning the model"	
	Experience and technical knowledge influence adherence to software configuration management models	" S4-2-P1: " (P1): Yes. And then, in the company I work at, we are working with AI now, and there, it is a step behind, but I am trying to implement a Git flow there. It is different, a little bit, one thing or another. Every company has more little differences, but I, more or less, are the ones who are doing it anyway. That's why I said that, for me, it's repetitive, but... Ok. (Moderator): And do you use the artifacts around the AI code in AI itself? (P1): There I am one step ahead. There I am forming the team first. First, I am convincing them to use Git. "S4-2-P1: "Ah, if you enter the job market you will discover... yeah... the pace here, you guys are very good. Oh, when we go to the market, the average is lower. So... there are other challenges, you know?"	S2-1-P1: "Here, which is an academic environment, I can't charge like in the corporate environment, right?"				
	Difficulty understanding branching model documentation			S1-5-P1: "Talking to you guys, yes." S1-5-P1: "Not 100%. But I read it once." S1-5-P4: "I had to reread it a little more, imagine the scenario."			

Machine Translated by Google

Theme	Code	Quotations GF1			Quotations GF2		
Branching model incompatibility with early or exploratory phases	Branching model is not suitable for early stage of project		S2-1-P4: "We didn't use it because we didn't make it available either, there was no stage or main, there was no stage, there was no production that would be the main, it was basically development and experiment"; S2-1-P4: "We weren't planning on making anything available"; S2-1-P1: "Because I had no idea that we would have to worry about these things. Maybe I also had no idea of the amount of people involved in the project. I had no idea about several things."; S2-1-P1: "I even suggest that this feedback should happen some time from now, it's not the best time now. Because even if we had started from the beginning like this... we would certainly have more contributions, but I think that now we start working on the product itself and this starts to make more sense."; S2-1-P4: "We hadn't really defined what the AI part would be like. Then, through the experiments, we were able to define the architecture that we showed."; S3-4-P4: "The impression I have today is that in the Review, right there, we present the results and from there some initiatives come out. So, I personally think that in the next Sprint Review we will deliver the code running, functioning, service, versioned and then start a process"		S4-1-P2: "I think that, maybe, we... we're parting a little bit, now In the beginning, mainly, because it's what I said, there are things that we just haven't done, you know that we're not at the right time to do them yet, like that. Maybe later on, we will be able to give better answers." S4-1-P3: "Yes. I think so, too, because as we... we said, we are very much in... in experimentation. So, as I said, in the first Sprint I didn't have much contact with versioning. But now, especially since we are working on the service, we will have more questions."	S1-2-P2: "The thing is, we don't have a homologation environment."; S1-2-P2: "our development is still kind of "flat" there in develop, we still don't... There's nothing in the homologation branch, and I think that's why it also, no... It hasn't come up yet, these conversations haven't come up"; S2-1-P2: "And I think this separation arises because I think we still don't... we're not even sure, either, about the things we want to pursue." S2-3-P2: "I think that in a more closed scenario, yes. It's just that, really, the moment may not be very favorable"; S2-3-P3: "Also. I think that's it for now. For the moment that we S1-1-cP3: "that we made these small changes to the two repositories is in the project." to meet the... The moment of the project."	S1-1-bP2: "I think that the fact that we split it into two repositories is very much part of the project's timing, like we don't even know which approach we are going to use overall, like that. So, besides being an AI scenario that is already unstable, it is a scenario where we are experimenting with different approaches because we don't know exactly what to use. So it is more unstable than normal. I think that in a "The moment we have a pipeline of everything we're going to do, then maybe we can put things together in a single repository and continue exactly as they are. But I think that due to the maturity of the project, we also have some of this difficulty" S1-1-cP2: "But I think that in a more controlled scenario it works, because I really think that our current scenario is very... It's very different, like that."
	Branching model is not suitable for exploratory or initial experimentation	S1-1-aP1: "When you do an experiment that is very out of the ordinary, you will do it in AI Experiments. But let's say we delivered a V1 of the MVP. For some reason, we need to do a V2. Got it? Go ahead. be apply this from here.			S4-1-P3: "Yes. I think so, also, because as we... we said, we are very much in... in experimentation. So, as I said, in the first I didn't have much contact with versioning in Sprint. But now, especially since we're working on the service, we're going to have more questions."		S1-1-aP2: "So, we were in this scenario where the things we generate are very unstable, and when we get to a result that we say, oh, this result was cool, we won't necessarily reuse that code. Sometimes, we'll start something else to have that structured, you know? Sometimes, experimentation is kind of... An experimentation. So, we... There is zero good practices. There is zero thought about becoming something here. (...) The goal, when you are experimenting, is not necessarily that it is good code to run, but rather that it is something that produces a result given a set of data." S1-1-bP2: "So, in addition to being an AI scenario that is already unstable, it is a scenario where we are experimenting with different approaches because we don't know exactly what to use. So it is more unstable than normal." S1-1-cP2: "I think because of the exploratory nature, like, I've worked on AI projects where we at least had a model type line that we were going to use, you know? And here we haven't even decided on that yet, so it's so... It's really exploration at a much higher level" S1-1-aP2: "We accepted that it was more unstable and that it needed more care, and this difference between training and inference, or sometimes even structured training, you know?"
Concerns about technical and human factors for adopting the branching model	Concern about scalability of the branching model		S2-1-P4: "If they decide to do an experiment together with the service, that's it. That's it and it can be continued. I just think that if we scale this up further, it will be a success."				
	Concern about privacy and security policies when using services external for data storage	S1-1-aP4: "We also have another problem, which is... Where are we going to control this... Where are we going to version this model. Because GitHub, GitHub doesn't have... It doesn't have its own structure, like, to... To version this model. What is most used in AI is Hugging Face. I don't know if you know. It's a GitHub for models and datasets. But, it's... Internally, it was kind of... Kind of confused about whether this will be used, whether it's safe for us to do things there. Because it's not on GitHub." " S1-1-aP1: " (P1): So, but that's what I'm saying. We don't have the infrastructure for that. Where am I going to store these models? (P4): The idea was on HuggingFace, but [X1] is having second thoughts. (P1): Where am I going to store the models and the dataset? With all this stuff that we have to sign.					
	Concern about challenges in using the branching model		S2-3-P1: "Yes, but I know the challenges that lie ahead, you know? So, I think that if the project is willing to deal with that, let's move forward."				
	Use of the branching model depends on team adoption		S2-1-P1: "We have to see how the team will receive this, it's very dependent on the team. A team can embrace it or the guy can say, "yeah.... I don't care". And then what are we going to do, you understand?"	S1-2-P1: "Man, everything you're saying there makes total sense. The question is whether it will be respected." S1-2-P1: "I have concerns about the motivational part." S1-2-P1: "I want to write code, do what you're talking about there, it's what I already do in my day to day life." S1-2-P1: "So, I see that here I'm going to do the same thing that I do, which is basically review code, guide, mentor, discuss."			
	System infrastructure is a limiting factor for some activities related to the branching model	S1-1-aP1: "But this experiment... You will version your code. Your setup. Yes. The result of the model is not to be versioned. But what they want is the traceability of your experiments. Of your code. changing. The parameters are changing." S1-1-aP1: "So one thing that confuses us is when you talk about an AI artifact. For me, an AI artifact is a model. And I don't version the model." S1-1-aP1: "(P1): The dataset is also not something that we version. (P4): But we need to. It's the same... We mentioned it at the beginning of the project Igual in HuggingFace. In HuggingFace, the dataset and models are versioned. Even like... Even he mentioned it. Did you have the teaching from code The code together with the dataset generated a model." S1-1-aP1: "So, let's go. How are you guys doing in relation to the other teams? Because today, what's the confusion? Our main, our production is on a single server. Do you understand? So I don't have a Quail 2 to run homologation, to release new features. So I have a server that runs prod, runs homolog and runs dev. In people's minds, this is confusing, you know? So we have some infrastructure challenges, which even for this to work nicely, you know? It has to be worked on, okay? There is a lot of coupling in these services. There is no clear bus that authenticates between one service and another. This is all due to back. Everything via Docker. So, there are some challenges that are not only for AI, but for other teams for you to implement." S1-1-aP1: "It's the same, I don't have a homologation server to run that stage there. I don't have a dev server to decouple it from the main. There are several challenges there." S1-1-aP1: "So, let's go. How are you guys doing in relation to the other teams? Because today, what's the confusion? Our main, our production is on a single server. Do you understand? So I don't have a Quail 2 there two to run homologation, to release new features. So I have a server that runs prod, runs homologation and runs dev. In the guys' heads, this becomes confusing, you know? So we have some infrastructure challenges, which even for this to work nicely, you know? It has to be worked on, okay? There is a lot of coupling in these services. There is no clear bus that authenticates between one service and another. This is all behind the scenes. Everything via Docker. So, there are some challenges that are not only for AI, but for other teams for you to implement." S1-1-aP1: "It's the same, I don't have a homologation server to run that stage there. I don't have a dev server to decouple it from the main. There are several challenges there." S1-1-cP1: " (P1): So, you here are [Moderator] who will navigate between the various teams. Fight there. So that we can actually have these two environments. (Moderator): Do you think that would be good? (P1): It would be good. With 20 people working here. (Moderator): If you have this infrastructure, would it be interesting? (P1): I think it's possible to run everything on the same server, doing a lot of stuff on the dockers. It's possible, but... I think it has to be done." S4-1-P1: "I think that the existence of the stage depends, in fact, on the existence of a homologation environment, if it doesn't make sense." S1-1-aP1: " (P1): So, but that's what I'm saying. We don't have the infrastructure for that. Where that I will keep these models? (P4): The idea was on HuggingFace, but [X1] is having second thoughts. (P1): Where am I going to store the models and the dataset? With all this stuff that we have to sign. "				
Hierarchical structure in technical decision making	Discussion on AI model approval with team leaders				S3-4-P2: I think... I think one thing is that it's easier to talk to people. team leaders. I don't know if this is a discussion that is worth extending to all developers, so. I think that's it, but it's not something that we have it very well defined, too, no."		
	Increased seniority for review on more stable branches		S2-1-P1: Will I restrict publishing on main? I'll probably have to restrict. I'm going to have to go in there and remove permission from... that I have today only one person can merge into main. Only one or two people can merge into stage. That's how the market is.	S1-2-P4: "Just... increasing the seniority level from develop to stage, stage to main.... It makes sense."			
Team collaboration and engagement	Decision-making is always carried out as a team		S2-2-P1: "No decision was made unilaterally. All of them were shared, there were people who gave their opinion, there were people who didn't, but it was always a team effort."	S1-2-P1: "I think the acceptance criteria there is the team's understanding." S1-2-P4: "So really, this is a team understanding."		S1-2-P2: "Naturally we look a lot at metrics related to the model itself"	
	Team commitment to reading documentation						S1-1-aP3: "I think everyone on the IA team read it because we had a day where we just sat down to go over the document."