Estimativa Atividades e Esforço  
Agente Coach Pro Ativo

**Projeto:** Agente de IA Proativo para Assistência em Tempo Real  
**Cliente:** Itaú Unibanco  
**Fornecedor:** Beyond Softworks  
**Data:** 27 de outubro de 2025  
**Prazo de Entrega:** 20 de novembro de 2025  
**Recursos Alocados:** 3 profissionais  
**Status:** Estimativa de atividades.

## DIVISÃO DE RESPONSABILIDADES

### Responsabilidades do Itaú

As seguintes atividades são de responsabilidade exclusiva do Itaú e **NÃO** fazem parte desta estimativa:

* Criar sigla APP para o agente
* Criar client\_id e secret\_path no vault para acesso ao IARA
* Criar repositório GitHub para o Agente
* Executar esteira do Agente para criar image\_uri
* Alterar repositório de infraestrutura para criar infra para o agente
* **Extrair conversas (dataset) do consórcio** ⚠️ *Dependência crítica*
* **Analisar melhores conteúdos** de acordo com as intenções selecionadas

### Responsabilidades da Beyond (Escopo desta Estimativa)

# Extrair intenções do dataset (teorema de pareto 80/20) - *Dependente do dataset do Itaú*

* **Desenvolver o agente pró-ativo** (core do projeto)
* **Criar processo de métricas**
* **Realizar testes e validação**

## PREMISSAS E CONDIÇÕES

### Premissas Técnicas Confirmadas

Com base nas discussões técnicas recentes, as seguintes incertezas foram **eliminadas**:

✅ **Azure AI Search:** Processo de clonagem e criação de bases de conhecimento está definido e validado  
✅ **Infraestrutura AWS:** SQS, ECS e Redis/Valkey serão provisionados pelo Itaú  
✅ **Acesso ao IARA:** Credenciais e configurações serão fornecidas pelo Itaú  
✅ **Framework:** LangChain será utilizado para desenvolvimento do agente  
✅ **Arquitetura:** Abordagem multi-agente está definida

### Dependências Críticas

Para que a Beyond possa iniciar o trabalho, as seguintes entregas do Itaú são **CRÍTICAS**:

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Responsável | Prazo Necessário | Impacto se Atrasar |
| **Dataset de conversas extraído** | Itaú | Até 30/Out | Bloqueia extração de intenções |
| Credenciais IARA (client\_id, secret\_path) | Itaú | Até 01/Nov | Atrasa integração com LLM |
| Infraestrutura AWS provisionada | Itaú | Até 04/Nov | Atrasa testes de integração |
| Repositório GitHub criado | Itaú | Até 28/Out | Atrasa início do desenvolvimento |

### Premissas de Execução

* **Equipe:** 3 recursos em tempo integral (Cientista de Dados + 2 Desenvolvedores)
* **Metodologia:** Entregas incrementais com validação contínua
* **Comunicação:** Alinhamentos diários com o Itaú para desbloqueio rápido
* **Escopo:** Congelado conforme definido neste documento
* **Qualidade:** Testes automatizados e validação de qualidade contínua

## DETALHAMENTO DE ATIVIDADES E ESFORÇOS

### ATIVIDADE 1: Extrair Intenções do Dataset (Pareto 80/20)

**Responsável:** Cientista de Dados  
**Dependência:** **Dataset de conversas do Itaú**  
**Esforço Estimado:** 20-28 horas  
**Duração:** 3-4 dias úteis

#### Descrição

Análise do dataset de conversas do domínio de consórcio para identificar e classificar as intenções dos clientes seguindo o princípio de Pareto (80/20), onde aproximadamente 80% das interações são cobertas por 20% das intenções.

#### Subtarefas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Análise exploratória** | Entender estrutura, volume, qualidade e padrões do dataset | 6-8 |
| **Identificação de intenções** | Aplicar técnicas de NLP (clustering, análise de frequência) para identificar intenções | 10-14 |
| **Classificação Pareto** | Classificar intenções em principais (80%) e cauda longa (20%) | 4-6 |

#### Entregáveis

* Taxonomia de intenções documentada
* Distribuição de frequência das intenções
* Mapeamento de intenções principais vs. cauda longa
* Recomendações de priorização

### ATIVIDADE 2: Desenvolver Agente Pró-Ativo

**Responsável:** 2 Desenvolvedores (1 Sênior + 1 Pleno)  
**Esforço Estimado:** 180-220 horas  
**Duração:** 15-17 dias úteis (paralelo)

Esta é a **atividade principal** do projeto, subdividida em módulos que serão desenvolvidos de forma incremental e paralela.

#### 2.1 Integrações Base

**Responsável:** Desenvolvedor Pleno  
**Esforço:** 28-36 horas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Integração com SQS** | Consumer de fila de entrada + Publisher de fila de saída | 12-16 |
| **Integração com Redis** | Cliente Redis para cache e gestão de estado de sessões | 10-12 |
| **Tratamento de erros** | Retry logic, Dead Letter Queue, Circuit Breaker | 6-8 |

**Entregáveis:** - Módulo de consumo de mensagens SQS funcional - Módulo de publicação de sugestões SQS funcional - Cliente Redis com operações de cache e sessão - Mecanismos de resiliência implementados

#### 2.2 Integração com IARA (LLM)

**Responsável:** Desenvolvedor Sênior  
**Esforço:** 32-40 horas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Cliente IARA** | Implementação de cliente HTTP para API IARA/Yara | 12-16 |
| **Engenharia de prompts** | Desenvolvimento de prompts para agentes (Triagem, Raciocínio, Validação) | 16-20 |
| **Gestão de tokens** | Controle de consumo de tokens e custos | 4-4 |

**Entregáveis:** - Cliente IARA funcional com suporte a múltiplos modelos - Biblioteca de prompts otimizados e versionados - Sistema de controle de tokens e custos

#### 2.3 Integração com Azure AI Search (RAG)

**Responsável:** Desenvolvedor Sênior + Desenvolvedor Pleno  
**Esforço:** 28-36 horas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Cliente Azure Search** | Implementação de busca vetorial e híbrida | 12-16 |
| **Otimização de queries** | Ajuste de parâmetros (top-K, threshold) e reranking | 10-12 |
| **Cache de buscas** | Cache inteligente de queries frequentes no Redis | 6-8 |

**Entregáveis:** - Módulo de RAG completo e funcional - Sistema de cache de buscas implementado - Estratégias de otimização documentadas

#### 2.4 Lógica do Agente Multi-Agente

**Responsável:** Desenvolvedor Sênior  
**Esforço:** 56-72 horas

Esta é a **parte mais complexa** do desenvolvimento, implementando a arquitetura multi-agente.

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Agente de Triagem** | Detecção de contexto suficiente, identificação de domínio, decisão de processamento | 20-24 |
| **Agente de Raciocínio** | Geração de sugestão usando LLM + RAG, personalização por contexto | 20-24 |
| **Agente de Validação** | Validação de relevância, segurança e consistência | 8-12 |
| **Orquestração** | Coordenação entre agentes, gestão de fluxo, prevenção de loops | 8-12 |

**Entregáveis:** - Agente de Triagem (Fast Reasoning) implementado - Agente de Raciocínio (Deep Reasoning) implementado - Agente de Validação (Quality Control) implementado - Orquestrador coordenando os agentes

#### 2.5 Testes e Validação

**Responsável:** Desenvolvedor Sênior + Desenvolvedor Pleno  
**Esforço:** 36-36 horas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Testes unitários** | Cobertura de componentes críticos (mínimo 70%) | 12-12 |
| **Testes de integração** | Validação de integrações (SQS, Redis, IARA, Azure Search) | 12-12 |
| **Testes end-to-end** | Validação de fluxo completo com cenários reais | 12-12 |

**Entregáveis:** - Suite de testes automatizados - Relatório de cobertura de testes - Validação de cenários críticos

### ATIVIDADE 3: Criar Processo de Métricas

**Responsável:** Cientista de Dados + Desenvolvedor Sênior  
**Esforço Estimado:** 32-40 horas  
**Duração:** 4-5 dias úteis (paralelo com desenvolvimento)

#### Descrição

Implementação de processo automatizado para validação da qualidade dos insights gerados pelo agente, conforme enfatizado nas discussões técnicas.

#### Subtarefas

|  |  |  |
| --- | --- | --- |
| Subtarefa | Descrição | Esforço (h) |
| **Definição de métricas** | Definir KPIs de qualidade, relevância, cobertura e experiência | 8-10 |
| **Implementação de coleta** | Instrumentar código para coletar métricas em tempo real | 12-16 |
| **Validação automática** | Implementar validação: insights vs. intenções e vs. KB | 8-10 |
| **Dashboard básico** | Criar visualização de métricas para acompanhamento | 4-4 |

#### Entregáveis

* Métricas de qualidade definidas e documentadas
* Sistema de coleta automatizada implementado
* Validação automática de insights funcionando
* Dashboard de acompanhamento de métricas

## RESUMO CONSOLIDADO

### Distribuição de Esforço por Atividade

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| # | Atividade | Esforço (h) | Recursos | Duração |
| **1** | **Extrair intenções (Pareto 80/20)** | 20-28 | 1 CD | 3-4 dias |
| **2** | **Desenvolver agente pró-ativo** | **180-220** | **2 Devs** | **15-17 dias** |
| 2.1 | → Integrações base (SQS, Redis) | 28-36 | 1 Dev Pleno | Semana 1-2 |
| 2.2 | → Integração IARA (LLM) | 32-40 | 1 Dev Sênior | Semana 1-2 |
| 2.3 | → Integração Azure Search (RAG) | 28-36 | 2 Devs | Semana 2 |
| 2.4 | → Lógica multi-agente | 56-72 | 1 Dev Sênior | Semana 2-3 |
| 2.5 | → Testes e validação | 36-36 | 2 Devs | Semana 3-4 |
| **3** | **Criar processo de métricas** | 32-40 | 1 CD + 1 Dev | 4-5 dias |
|  | **TOTAL GERAL** | **232-288** | **3 recursos** | **~3,5 semanas** |

### 

### Análise de Viabilidade

|  |  |  |
| --- | --- | --- |
| Métrica | Valor | Observação |
| **Esforço Total Estimado** | 232-288 horas | Cenário otimista-pessimista |
| **Dias Úteis Disponíveis** | 17 dias | 28/Out a 20/Nov (descontando feriados) |
| **Capacidade com 3 Recursos** | 408 horas | 3 × 17 dias × 8h/dia |
| **Margem de Segurança** | 120-176 horas | 29-43% de folga |
| **Viabilidade** | **VIÁVEL** | Com boa margem para imprevistos |

## CRONOGRAMA MACRO

### Semana 1: 28/Out - 01/Nov (5 dias úteis)

**Dependência crítica:** Receber dataset do Itaú até 30/Out

|  |  |  |
| --- | --- | --- |
| Recurso | Atividades | Horas |
| **Cientista de Dados** | Análise do dataset + Extração de intenções | 20-28 |
| **Dev Sênior** | Setup + Cliente IARA + Prompts iniciais | 32-40 |
| **Dev Pleno** | Setup + Integrações SQS/Redis | 28-36 |

**Entregas da Semana:** - Taxonomia de intenções (80/20) documentada - Cliente IARA funcional - Integrações SQS/Redis funcionais

### Semana 2: 04/Nov - 08/Nov (5 dias úteis)

|  |  |  |
| --- | --- | --- |
| Recurso | Atividades | Horas |
| **Cientista de Dados** | Processo de métricas (definição + coleta) | 20-26 |
| **Dev Sênior** | Integração Azure Search + Agente de Triagem | 40-52 |
| **Dev Pleno** | Suporte Azure Search + Cache | 28-36 |

**Entregas da Semana:** - Integração RAG funcional - Agente de Triagem implementado - Processo de métricas (parcial)

### Semana 3: 11/Nov - 15/Nov (5 dias úteis)

**Feriado:** 15/Nov (Proclamação da República) - **4 dias úteis efetivos**

|  |  |  |
| --- | --- | --- |
| Recurso | Atividades | Horas |
| **Cientista de Dados** | Validação automática + Dashboard | 12-14 |
| **Dev Sênior** | Agente de Raciocínio + Validação + Orquestração | 56-72 |
| **Dev Pleno** | Suporte desenvolvimento + Testes unitários | 32-32 |

**Entregas da Semana:** - Agente completo (Triagem + Raciocínio + Validação) - Orquestração multi-agente funcionando - Processo de métricas completo

### Semana 4: 18/Nov - 20/Nov (3 dias úteis)

**Finalização e validação**

|  |  |  |
| --- | --- | --- |
| Recurso | Atividades | Horas |
| **Cientista de Dados** | Validação de métricas + Ajustes finais | 8-8 |
| **Dev Sênior** | Testes de integração + Ajustes | 12-12 |
| **Dev Pleno** | Testes E2E + Documentação | 12-12 |

**Entregas da Semana:** - Solução completa testada - Documentação técnica - Pronta para deploy

## RISCOS E MITIGAÇÕES

### Riscos Técnicos

|  |  |  |  |
| --- | --- | --- | --- |
| Risco | Prob. | Impacto | Mitigação |
| **Dataset não entregue no prazo** | Média | Crítico | Comunicação diária com Itaú; ter dataset de backup para testes |
| **Latência do LLM acima do aceitável** | Alta | Alto | Testar modelos rápidos logo na semana 1; otimizar prompts |
| **Complexidade da orquestração maior que esperado** | Média | Médio | Começar com versão simplificada; iterar |
| **Bloqueios por credenciais/acessos** | Média | Médio | Solicitar tudo antecipadamente; ter tarefas alternativas |

### Riscos de Gestão

|  |  |  |  |
| --- | --- | --- | --- |
| Risco | Prob. | Impacto | Mitigação |
| **Mudanças de escopo** | Alta | Alto | Congelar escopo; mudanças vão para fase 2 |
| **Indisponibilidade de recursos** | Baixa | Alto | Ter backup de recursos; documentar conhecimento |
| **Feriado 15/Nov** | Certa | Médio | Já considerado no cronograma (4 dias úteis na semana 3) |

## ENTREGÁVEIS FINAIS

Ao final do projeto (20/Nov/2025), a Beyond entregará:

### 1. Agente Pró-Ativo Funcional

* Código-fonte completo e documentado
* Arquitetura multi-agente implementada (Triagem, Raciocínio, Validação)
* Integrações funcionais:
  + SQS (consumer e publisher)
  + Redis (cache e gestão de estado)
  + IARA (modelos de linguagem)
  + Azure AI Search (RAG)
* Testes automatizados (unitários, integração, E2E)
* Cobertura de testes mínima de 70%

### 2. Análise de Intenções

* Taxonomia de intenções (Pareto 80/20)
* Distribuição de frequência
* Mapeamento de intenções principais vs. cauda longa
* Recomendações de priorização

### 3. Processo de Métricas

* Métricas de qualidade definidas
* Sistema de coleta automatizada implementado
* Validação automática de insights (vs. intenções e vs. KB)
* Dashboard de acompanhamento

### 4. Documentação Técnica

* README com instruções de setup e execução
* Documentação de arquitetura e componentes
* Documentação de APIs e integrações
* Guia de troubleshooting básico
* Exemplos de uso

## CONDIÇÕES PARA INÍCIO

Para que a Beyond possa iniciar o trabalho conforme planejado, as seguintes condições **DEVEM** ser atendidas:

### Até 29/Out (Quarta-feira)

* ✅ Repositório GitHub criado e acessos concedidos
* ✅ Confirmação de alocação dos 3 recursos

### Até 30/Out (Quarta-feira)

* ✅ **Dataset de conversas do consórcio disponibilizado** (CRÍTICO)
* ✅ Documentação de acesso ao dataset

### Até 01/Nov (Sexta-feira)

* ✅ Credenciais IARA (client\_id e secret\_path)
* ✅ Documentação da API IARA

### Até 04/Nov (Segunda-feira)

* ✅ Infraestrutura AWS provisionada (SQS, Redis, ECS)
* ✅ Credenciais de acesso ao Azure AI Search
* ✅ Documentação de acesso aos ambientes

## OBSERVAÇÕES IMPORTANTES

### Escopo Congelado

O escopo definido neste documento está **congelado**. Qualquer solicitação de mudança ou adição de funcionalidades durante a execução será tratada como **Fase 2** (pós 20/Nov).

### Comunicação Contínua

Será estabelecida comunicação diária entre Beyond e Itaú para: - Acompanhamento de progresso - Desbloqueio rápido de impedimentos - Validação de entregas incrementais - Alinhamento de expectativas

### Qualidade vs. Prazo

A estimativa prioriza a entrega no prazo (20/Nov) mantendo qualidade mínima aceitável: - Testes automatizados (70% de cobertura) - Validação de cenários críticos - Documentação essencial

Refinamentos e otimizações adicionais podem ser tratados em fases posteriores.

## CONCLUSÃO

A estimativa preliminar apresentada demonstra que é **VIÁVEL** entregar o Agente Coach Pro Ativo até **20 de novembro de 2025** com **3 recursos** da Beyond, desde que:

1. ✅ As **dependências críticas** do Itaú sejam atendidas nos prazos definidos
2. ✅ O **escopo seja mantido** conforme este documento
3. ✅ Haja **comunicação diária** e desbloqueio rápido de impedimentos
4. ✅ Os **3 recursos** estejam alocados em tempo integral

O esforço total estimado de **232-288 horas** cabe confortavelmente nas **408 horas disponíveis**, deixando uma margem de segurança de **29-43%** para imprevistos, ajustes e refinamentos.

A eliminação de incertezas sobre o Azure AI Search e a clareza na divisão de responsabilidades aumentam significativamente a confiança nesta estimativa.

**Elaborado por:** Guilherme Cabreira  
**Data:** 27 de outubro de 2025  
**Versão:** 1.0 (Preliminar)  
**Próxima Revisão:** Após recebimento do dataset (30/Out)