



Junho 13, 2025

Mineração de Dados - NutriBot

Turma 24-25

Índice

1)	Introdução e motivação	5
2)	Objetivos	6
3)	Role A	7
3.1)	Cronograma	7
3.2)	Front-end	7
3.2.1)	Interfaces	8
3.3)	Resultados	10
3.3.1)	Primeira fase de Avaliação	10
3.3.2)	Segunda fase de Avaliação	11
4)	Role B	13
4.1)	Arquitetura Inicial	13
4.2)	Arquitetura Final	14
4.2.1)	Global Agent	14
4.2.2)	Bases de Dados	15
4.2.3)	Rotas e Fluxo	16
5)	Role C	18
5.1)	Arquitetura Geral do Sistema	18
5.2)	Metodologia de Extração de Dados	19
5.3)	Processamento e Transformação da Informação	19
5.4)	Estruturação dos Dados	20
5.5)	Conclusões e Trabalho Futuro	21
6)	Role D	22
6.1)	Entidades	22
6.1.1)	User	22
6.1.2)	Conversation	22
6.1.3)	Message	23
6.1.4)	UserInfo	23
6.2)	Limitações	25
7)	Role E	26
7.1)	<i>Workflow</i>	26
7.1.1)	<i>Payload</i> Inicial	26
7.1.2)	RAG (Inserção e Vetorização)	27
7.1.3)	RAG (Recolha de Dados)	27
7.1.4)	Gerar <i>Prompt</i> Final	28
7.1.5)	Enviar Pergunta para o LLM	29
7.1.6)	Retornar Resposta	29
7.2)	Configuração	29
7.3)	<i>Benchmarking</i>	29
8)	Especialista em Nutrição	31
8.1)	Análise do tema e das fontes de dados relevantes	31
8.2)	Metodologia de Recolha e Processamento de Dados	31
8.2.1)	Estratégia de Web Scraping	31
8.2.2)	Ferramentas de Extração	31
8.2.3)	Utilização de queries	32
8.2.4)	Definição de Keywords por Domínio	32
8.3)	Avaliação do pipeline RAG	32

8.3.1)	Objetivos do Teste	32
8.3.2)	Questão e Resultado Obtido	32
8.3.3)	Recuperação	33
8.3.4)	Geração	33
8.3.5)	Observações Técnicas	33
8.3.6)	Avaliação Quantitativa	34
8.4)	<i>Benchmarks</i> de Modelos	34
8.4.1)	Avaliação da Qualidade de Resposta dos Modelos	34
8.4.2)	Modelo Selecionado	35
8.5)	Configuração do Agente	36
9)	Especialista em Suplementação e Fármacos	37
9.1)	Análise do tema e das fontes de dados	37
9.2)	Metodologia de Recolha e Processamento de Dados	38
9.2.1)	Extração dos dados	38
9.2.2)	Processamento dos dados	38
9.2.3)	Armazenamento dos dados	39
9.3)	<i>Benchmarks</i>	39
9.3.1)	Avaliação da Qualidade de Resposta dos Modelos	40
9.3.2)	Modelo Selecionado	43
9.4)	Configuração do Agente	43
10)	Especialista em Atividade Física	44
10.1)	Análise do Tema e das Fontes de Dados Relevantes	44
10.2)	Metodologia de Recolha e Processamento de Dados	44
10.2.1)	Estratégia de Web Scraping	44
10.2.2)	Ferramentas de Extração	45
10.2.3)	Definição de Keywords por Domínio	45
10.2.4)	Chunking de Dados	46
10.2.5)	Hierarquia	47
10.3)	<i>Benchmarks</i> de Modelos	47
10.3.1)	Avaliação da Qualidade de Resposta dos Modelos	47
10.3.2)	Modelo Selecionado	48
10.4)	Configuração do Agente	48
11)	Especialista em Hábitos Diários	50
11.1)	Análise do tema e das fontes de dados relevantes	50
11.2)	Extração de Conhecimento Externo	50
11.2.1)	Extração Manual	50
11.2.2)	Recolha Automatizada de Dados	50
11.2.3)	Processamento de Dados	51
11.2.4)	Chunking de Dados	51
11.3)	<i>Benchmarks</i>	51
11.3.1)	Avaliação da Qualidade de Resposta dos Modelos	51
11.3.2)	Tempo de Execução	54
11.3.3)	Modelo Escolhido	54
11.4)	Configuração do Agente	54
12)	Especialista em Check-ups/acompanhamento médico de medicina preventiva	55
12.1)	Exploração do Tema e Identificação de Fontes Relevantes	55
12.2)	Extração de Conhecimento Externo	55
12.2.1)	Extração Manual de Informação Confiável	55

12.2.2)	Recolha Automática de Dados	55
12.2.3)	Tratamento e Preparação dos Dados Recolhidos	55
12.2.4)	Chunking de Dados	56
12.3)	Context Prompt Usado	56
12.4)	Benchmarks	56
12.4.1)	Avaliação de Qualidade: Exemplos Qualitativos	56
12.4.2)	Tabela de Avaliação por IA e Pergunta	59
12.4.3)	Modelo Escolhido	62
13)	Conclusão	63
	Bibliografia	64

1) Introdução e motivação

No panorama atual, marcado por um acesso cada vez mais amplo e imediato à informação sobre saúde e bem-estar, torna-se essencial assegurar que essa informação seja não apenas fiável, mas também acessível, contextualizada e personalizada às necessidades de cada indivíduo. Apesar da abundância de conteúdos disponíveis online, muitos utilizadores continuam a enfrentar dificuldades na identificação de fontes credíveis, na interpretação de recomendações científicas e na aplicação prática desses conhecimentos no seu quotidiano. Esta lacuna abre espaço para desinformação, escolhas inadequadas e até riscos para a saúde.

Neste enquadramento, o presente projeto propõe o desenvolvimento de um *chatbot* interativo orientado para a promoção de um estilo de vida saudável, integrando conceitos e metodologias abordadas na unidade curricular de Mineração de Dados. Esta solução visa mitigar as barreiras ao acesso a informação de qualidade, promovendo decisões informadas e comportamentos mais saudáveis.

O principal objetivo do mesmo é fornecer recomendações práticas, adaptadas ao perfil do utilizador, sobre temas como atividade física, nutrição, suplementos alimentares e fármacos, hábitos de vida saudáveis e a realização regular de *check-ups* médicos. Para tal, o sistema baseia-se na recolha, tratamento e análise de dados provenientes de fontes científicas e institucionais credíveis, aplicando técnicas avançadas de pré-processamento, extração de conhecimento e integração de dados heterogéneos.

Adicionalmente, foi dada especial atenção à estruturação das respostas geradas pelo agente conversacional, com o intuito de assegurar uma comunicação fluida, clara e útil, capaz de estabelecer uma relação de confiança com o utilizador. Neste processo, foram exploradas e testadas diferentes abordagens linguísticas e semânticas, de modo a otimizar a qualidade e a pertinência das interações.

Este relatório encontra-se estruturado de forma a refletir de modo claro as várias dimensões do trabalho desenvolvido. Inicialmente, são apresentados os objetivos gerais e específicos do projeto. Seguidamente, descreve-se detalhadamente o contributo de cada *Role* (A, B, C, D e E), incluindo as tarefas atribuídas e as tecnologias envolvidas. É ainda dedicada especial atenção ao trabalho realizado por cada grupo, com destaque para as diferenças observadas entre as *roles* C e E, diferenças essas que, embora partilhassem domínios semelhantes, se distinguiram ao nível das abordagens técnicas, prioridades e métodos de integração.

Por fim, são discutidas as principais dificuldades encontradas ao longo do desenvolvimento, as soluções adotadas e as aprendizagens obtidas, culminando numa avaliação crítica do resultado final.

Com este projeto, pretende-se demonstrar não apenas a aplicabilidade prática das técnicas de mineração de dados em contextos reais, mas também a sua relevância social enquanto ferramenta de apoio à literacia em saúde, contribuindo para um acesso mais equitativo, fiável e personalizado à informação.

2) Objetivos

Os objetivos do projeto foram separados em *Roles* para organizar o desenvolvimento geral. A *Role A* teve o papel de coordenador geral do projeto, realizando também o *frontend*, a *Role B* focou-se na definição e implementação do sistema multiagente, a *Role C* dedicou-se ao tratamento e organização dos dados a utilizar, a *Role D* centrou-se na recolha e gestão da informação dos utilizadores e por fim, a *Role E* teve por foco os modelos do sistema, desde a sua definição até à sua validação.

- **Role A**
 - Desenvolvimento do *frontend*.
 - Cumprimento dos prazos estabelecidos.
 - Comunicação eficaz entre os diferentes grupos.
- **Role B**
 - Identificação das entidades fundamentais do sistema.
 - Definição e implementação da arquitetura geral do sistema.
 - Escolha das tecnologias a utilizar.
- **Role C**
 - Definição das fontes de dados.
 - Seleção e organização dos dados.
 - Extração de conhecimento das fontes.
 - Resolução de divergências entre fontes.
 - Normalização das entidades.
 - Realização de pesquisa na Internet.
- **Role D**
 - Definição do método de recolha de informação do utilizador.
 - Escolha da informação relevante do utilizador.
 - Estratégia de armazenamento da informação do utilizador.
 - Definição das tecnologias a utilizar.
 - Implementação de recolha dinâmica da informação do utilizador.
- **Role E**
 - Prompt engineering.
 - Escolha do modelo.
 - Desenho do workflow de processamento.
 - Treino do modelo.
 - Validação do modelo.
 - Priorização de fontes confiáveis.
 - Validação automática do modelo.

3) Role A

A *Role A* é responsável por garantir que os objetivos do grupo sejam atingidos de forma eficiente e alinhada com a visão global do projeto. Atua como elo central entre os diferentes grupos de trabalho, coordenando o desenvolvimento e a integração da aplicação em escala global. Além disso, lidera a gestão do *front-end*, assegurando consistência na experiência do utilizador, cumprimento com os requisitos técnicos e colaboração contínua entre membros de uma *Role* ou dentro do próprio grupo.

3.1) Cronograma

Considerando os objetivos levantados pelo coletivo, desenvolvemos um cronograma, sobre a forma de diagrama de Gantt, para garantir a sua conclusão.

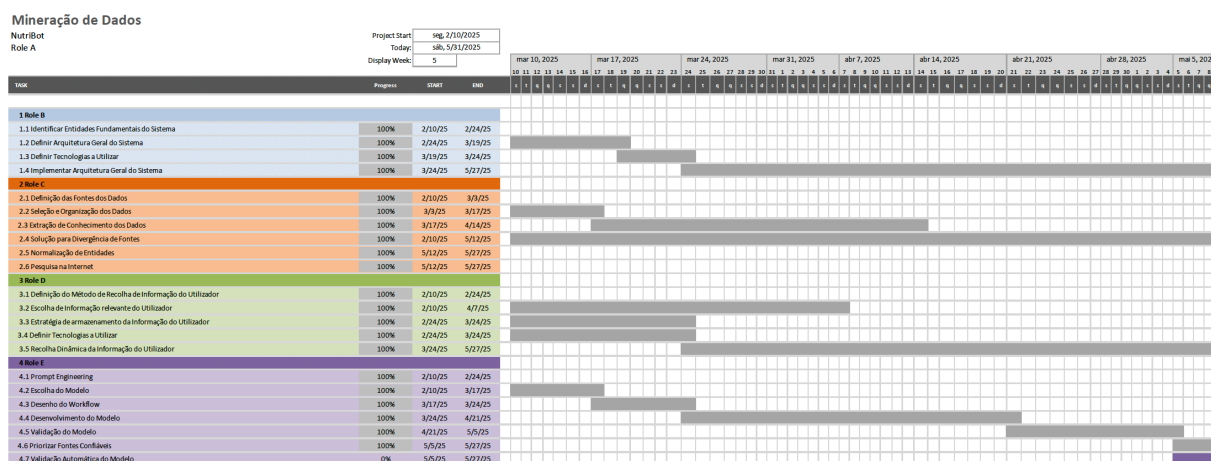


Figura 1: Diagrama de Gantt

De modo geral, os grupos e os respetivos *roles* conseguiram cumprir o agendamento previsto e concluir os objetivos definidos. No entanto, importa referir que se verificou um ligeiro atraso de aproximadamente três dias, o qual se atribui, em grande parte, à heterogeneidade entre os elementos das equipas. Acresce ainda que, por ausência de alternativas viáveis, a tarefa de “Validação Automática do Modelo” ficou por concluir, uma vez que não foi possível encontrar uma solução *open-source* ou gratuita que permitisse a sua implementação de forma eficaz.

3.2) Front-end

A conceção do *front-end* ficou da responsabilidade da *Role A*. Nesta componente, tivemos como objetivo entregar interfaces que demonstrassem o trabalho desenvolvido, como também aliar o projeto a uma das plataformas mais utilizadas na indústria, mais concretamente, o *web development*. Assim, optamos por utilizar a *framework Next.js* para o desenvolvimento desta camada, visto que habilita possibilidades como renderização a nível do servidor, e também é altamente popular, prática e simples. Ainda, optamos por tomar conta da lógica de autenticação da plataforma, de modo a provar a interoperabilidade do sistema desenhado.

Primeiramente, optamos por modelar como seria a travessia entre páginas dentro da interface. Para esse efeito, recorreremos a um diagrama de *webflow*.

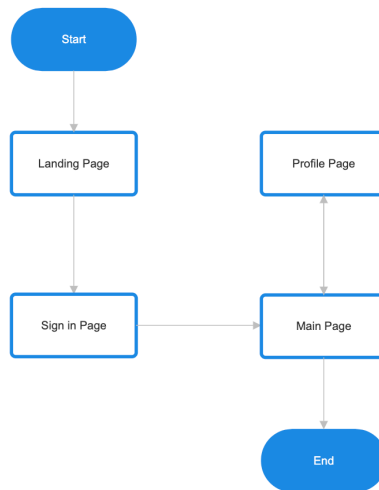


Figura 2: Diagrama de *Webflow*

Na figura Figura 2, observamos um fluxo esperado para um novo utilizador no seu primeiro início na plataforma, onde é esperado que o processo inicie na *Landing page*, onde terá acesso a algumas informações da plataforma, siga para a autenticação, proceda à interação com os agentes via *chat* e para a adição de informação sobre o seu perfil.

3.2.1) Interfaces

Na figura Figura 3 encontramos a primeira interface, a *Landing page* do projeto. Aqui é dada a opção ao utilizador para iniciar sessão, podendo aceder à plataforma, como também são expostas algumas informações à cerca do projeto desenvolvido.

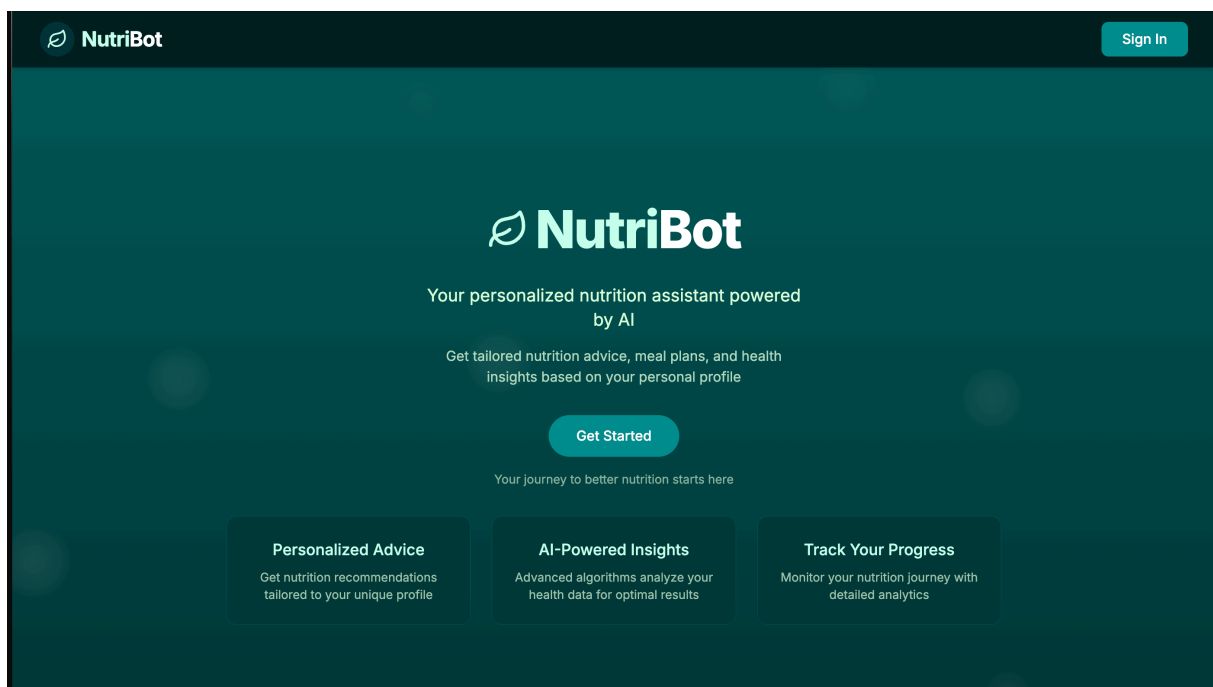


Figura 3: Landing page

De seguida, na figura Figura 4, encontramos uma interface disponibilizada pela *API* da Clerk, que é a plataforma externa utilizada para gerir a autenticação da nossa aplicação. É de notar que o uso desta *API* em nada alterou a conceptualização feita, tanto pela *Role B*, como também com a *Role D*.

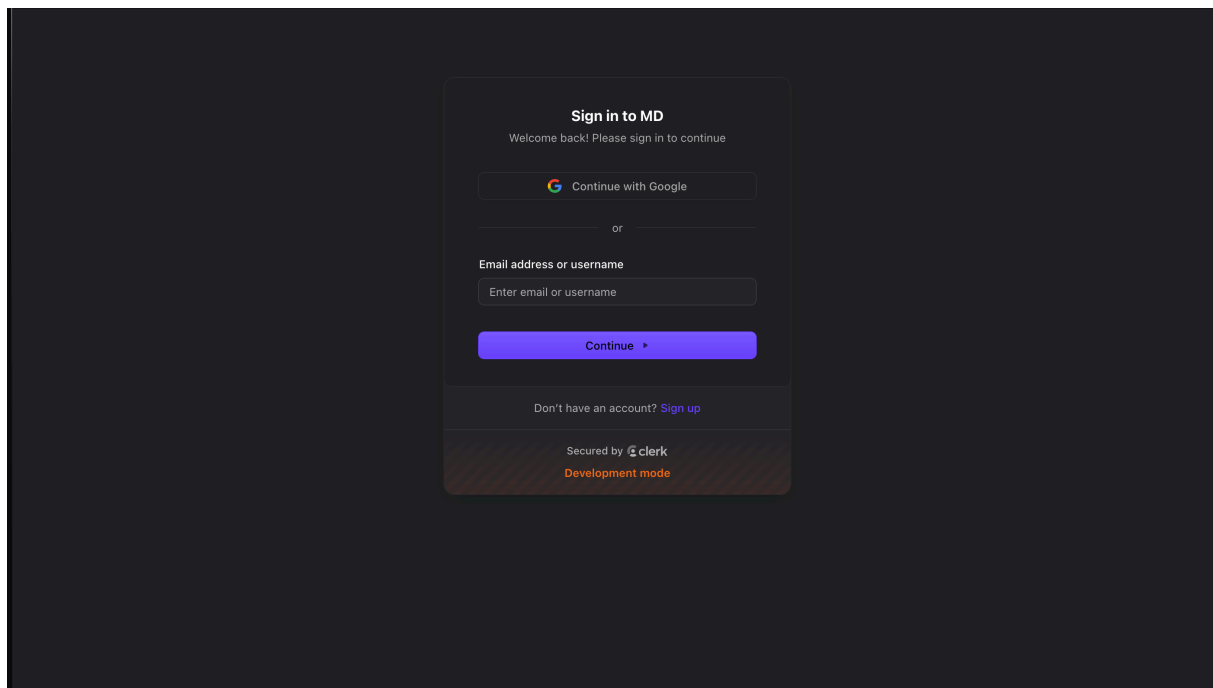


Figura 4: *Sign-in page*

A seguir, encontramos a página principal da aplicação. Nesta página, um utilizador pode criar uma conversa, interagir com uma conversa, trocar entre conversas já existentes, aceder ao seu perfil, aceder às definições da conta, tais como, apagar conta ou proceder ao *Sign-out*

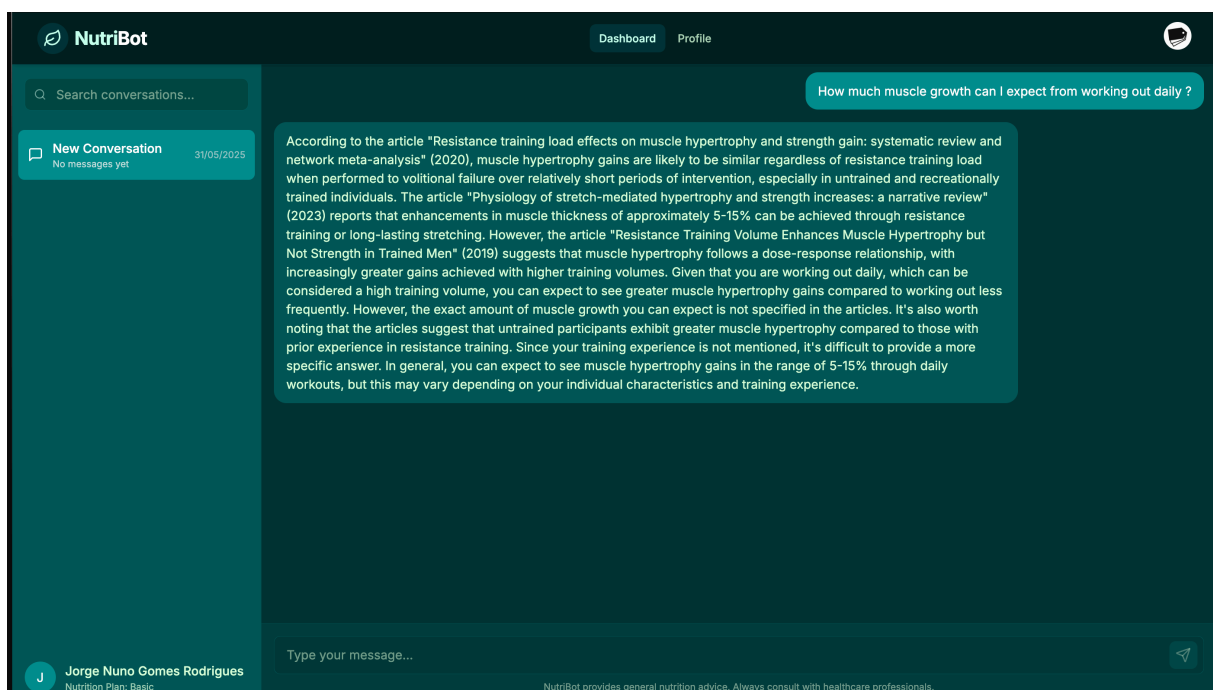


Figura 5: *Main page*

Para terminar, segue a página de perfil, onde é esperado que os utilizadores forneçam informações que apoiem a geração dos agentes. Qualquer informação fornecida poderá ser alterada pelo utilizador. Aqui são encontrados todos os campos levantados pela *Role D*.

Figura 6: *Profile page*

3.3) Resultados

A responsabilidade pela realização dos testes à plataforma coube à *Role A*. Para conduzir esta avaliação, foi adotado um método de caixa preta, procurando simular o comportamento típico de um utilizador da aplicação. A metodologia de avaliação seguiu, em parte, o formato sugerido pela *Role E*, incluindo também a componente que se ocupa de redirecionar as questões para os agentes especializados. O processo de avaliação foi dividido em duas fases distintas, nas quais foram colocadas perguntas imparciais direcionadas a cada um dos temas abordados pelos agentes.

3.3.1) Primeira fase de Avaliação

A primeira fase foi concebida com o intuito de servir como molde para que os grupos identificassem falhas nas suas componentes, razão pela qual as perguntas foram construídas de forma imparcial. Os resultados obtidos nesta fase, para um total de 50 perguntas (10 dirigidas a cada agente), foram registados numa tabela de resultados, que permitiu evidenciar pontos fortes e aspetos a melhorar na aplicação. Esta etapa foi essencial para o diagnóstico inicial do desempenho da plataforma e orientou as primeiras correções.

	Perguntas Encaminhadas Corretamente	Perguntas Encaminhadas Incorretamente	Taxa de sucesso
Global	50	0	100%

Tabela 1: Resultados do primeiro Benchmark (agente global)

Verificamos que o agente global tem uma taxa de sucesso de 100%, encaminhando todas as perguntas para os agentes corretos.

Grupo	Rating mais Alto	Rating mais Baixo	Avaliação Média
Grupo 1	9.0	7.0	8.5
Grupo 2	8.5	6.0	7.1
Grupo 3	8.0	2.0	5.6
Grupo 4	9.5	7.0	8.1
Grupo 5	8.0	4.0	6.3
Geral	9.5	2.0	7.1

Tabela 2: Resultados do primeiro Benchmark (agentes individuais)

Ao analisar os nossos resultados, verificamos que todos os agentes receberam uma avaliação positiva, apesar de haver alguma heterogeneidade entre a avaliação média de cada agente. Existe a possibilidade de isto se verificar por causa de diferenças entre cada um dos agentes, mas o facto de realizarmos uma quantidade baixa de perguntas a cada agente (10) poderá ter influenciado os resultados, inflacionando alguns e diminuindo outros comparativamente à performance real do agente. Assim, na próxima avaliação, será importante aumentar o número de perguntas feitas a cada agente.

3.3.2) Segunda fase de Avaliação

A segunda fase serviu para aferir o estado final da plataforma após as melhorias implementadas. Os testes realizados procuraram verificar a robustez do sistema num contexto mais próximo do uso real. Assim como na primeira fase, os resultados foram organizados numa tabela, acompanhada por uma análise qualitativa dos mesmos. Esta análise permitiu verificar a evolução da aplicação ao longo do projeto, demonstrando a maturidade alcançada pela solução desenvolvida. Foram realizadas 100 perguntas, com o intuito de serem feitas 20 a cada agente. Como, ao contrário da fase anterior, algumas perguntas foram dirigidas a um agente diferente do suposto, apresentamos também a quantidade de perguntas feitas a cada agente.

	Perguntas Encaminhadas Corretamente	Perguntas Encaminhadas Incorretamente	Taxa de sucesso
Global	96	4	96%

Tabela 3: Resultados do segundo Benchmark (agente global)

Podemos verificar que, nesta fase, ao contrário da anterior, o agente global encaminhou algumas perguntas para os agentes errados. Analisando estes erros, observamos que foram todas as perguntas encaminhadas ao Agente 5 que foram erradamente encaminhadas para o Agente 4, pois englobavam tópicos transversais a ambos os agentes, como, por exemplo, hábitos para prevenção de doenças.

Grupo	Rating mais Alto	Rating mais Baixo	Avaliação Média	Número de Perguntas
Grupo 1	9.0	3.0	8.0	20
Grupo 2	9.0	7.0	8.12	20
Grupo 3	9.5	6.0	8.57	20
Grupo 4	9.5	6.0	7.90	24
Grupo 5	10.0	6.5	8.78	16
Geral	10.0	3.0	8.24	100

Tabela 4: Resultados do segundo Benchmark (agentes individuais)

Analisando os resultados da segunda avaliação, conseguimos verificar uma melhoria na avaliação média geral, apesar de alguns agentes descerem na sua avaliação média comparativamente à primeira. Estas diferenças podem ser atribuídas às melhorias feitas a cada agente, mas também devem-se ao facto de o número de perguntas ter duplicado relativamente à avaliação anterior. Isto pode ser verificado pelo facto das avaliações médias de cada agente serem mais próximos umas das outras comparativamente à avaliação anterior, mostrando que, com uma maior quantidade de perguntas, os agentes têm uma performance mais similar. Por exemplo, o agente com a avaliação média mais desviada da média (Grupo 5) é o que elaborou o menor número de respostas.

Dos resultados, podemos tirar algumas conclusões e pontos onde seria possível melhorar.

Em termos do *benchmarking* em si, verificamos que, realizando mais perguntas, a avaliação do modelo era mais próxima da realidade. Portanto, fazer um *benchmarking* com uma quantidade ainda mais elevada de perguntas produziria resultados ainda mais fidedignos. Também deveria ser explorado outro método de avaliar as respostas sem utilizar somente a avaliação por modelos externos, como por exemplo uma avaliação manual ou com base em respostas definidas anteriormente, mas isto necessitaria de uma equipa com o conhecimento médico necessário para realizar estas tarefas.

Em termos do sistema em si, verificamos como, por vezes, perguntas contêm informação transversal a vários agentes. Com o nosso sistema atual de encaminhamento, o agente global tem de decidir o agente a que vai enviar a pergunta. Se implementássemos uma fusão de respostas de diferentes agentes, poderíamos encaminhar a pergunta a qualquer agente que fosse relevante e depois juntar a informação toda numa única resposta, resolvendo assim qualquer problema de encaminhamento. Também é possível ainda verificar uma heterogeneidade entre os agentes, especialmente em relação à formatação das respostas de cada agente, com, por exemplo, alguns respondendo com a formatação em Markdown. O tempo que cada agente demora a responder a uma pergunta também é demasiado variado entre os agentes, alguns demorando pouco tempo para gerar a resposta e outros mais. Todas estas diferenças deveriam ser minimizadas o máximo possível, de forma a oferecer uma experiência mais consistente aos utilizadores.

4) Role B

A Role B foi responsável pelo desenho da arquitetura global do sistema multiagente, assegurando a integração entre os diferentes agentes especializados. Desenvolveu-se um agente orquestrador central (“Global Agent”) e uma API (Application Programming Interface) modular que garante a interoperabilidade, escalabilidade e coerência na comunicação entre os diversos componentes do sistema.

4.1) Arquitetura Inicial

A arquitetura inicialmente proposta tinha como objetivo modularizar ao máximo as responsabilidades do sistema multiagente, garantindo flexibilidade, escalabilidade e visibilidade sobre o seu funcionamento interno. Foram idealizados diversos componentes com papéis distintos, nomeadamente o “MetaAgent” (LLM Engine), um “DataLake”, uma *Dashboard* de monitorização com sistema de *logs* e uma estrutura onde o “Orchestrator” desempenhava um papel ativo na comunicação com os agentes especialistas.

Contudo, durante o desenvolvimento, várias decisões técnicas levaram à simplificação do sistema. O “MetaAgent”, responsável por processar perguntas através de um modelo LLM genérico antes de contactar os agentes, foi removido. Considerou-se mais direto e eficaz confiar apenas nas capacidades de cada agente especializado para processar essas perguntas. O “DataLake”, apesar de ser uma proposta interessante para centralização e reproprocessamento dos dados, revelou-se desnecessário face à utilização de uma base de dados documental (*MongoDB*), mais leve e adequada ao contexto do projeto.

As ferramentas planeadas para monitorização do sistema, a *Dashboard* e os seus *logs*, não aparecem representadas na arquitetura final, embora tenham sido parcialmente desenvolvidas para apoio durante a implementação. Quanto ao “Orchestrator”, a sua função inicial de gerir diretamente a comunicação com os agentes foi repensada. Na prática, é a API que comunica com os agentes especialistas após receber o nome do agente indicado pelo Global Agent, o qual mantém apenas o registo do último agente utilizado para garantir consistência nas interações subsequentes.

Esta evolução reflete uma abordagem mais pragmática e orientada à funcionalidade, onde a complexidade arquitetural foi reduzida em prol da clareza, eficiência e rapidez de desenvolvimento.

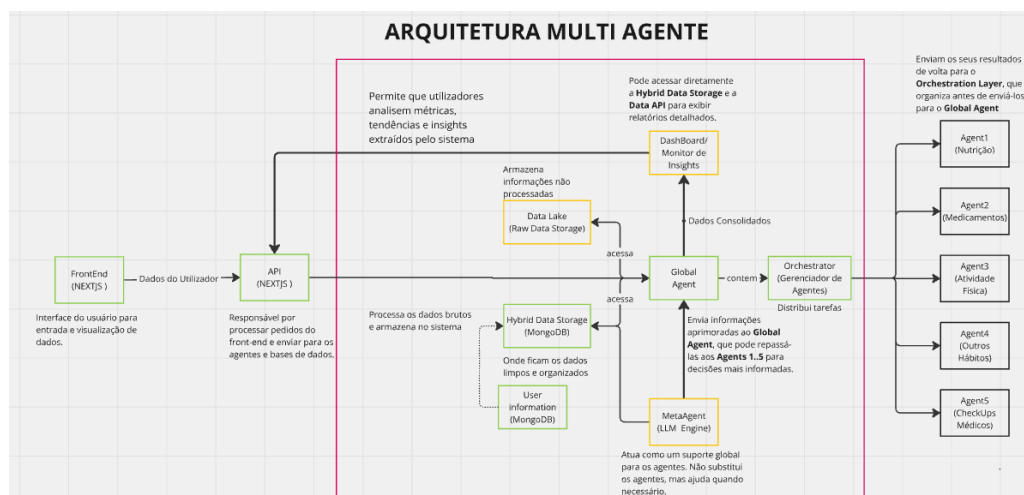


Figura 7: Arquitetura Inicial Idealizada

4.2) Arquitetura Final

Na arquitetura final do sistema multiagente, o foco passou a ser a eficiência na comunicação entre os componentes, a modularidade da API e a centralização da lógica de decisão no Global Agent, refletindo uma abordagem mais simplificada e pragmática face à proposta inicial.

O sistema é composto por uma interface *front-end* desenvolvida em *Next.js*, que interage com a API Gateway, também implementada em *Next.js*, responsável por encaminhar os pedidos do utilizador. Esta, por sua vez, comunica com uma API construída em *Flask*, uma *micro-framework* escrita em *Python*, que coordena o acesso à base de dados *MongoDB* e o envio de perguntas aos agentes especializados.

A componente central desta arquitetura é o Global Agent, que determina qual o agente mais apropriado para responder a cada pergunta, com base na primeira interação da conversa. O nome do agente selecionado é armazenado numa base de dados, em conjunto com outros dados, garantindo consistência nas interações seguintes. Esta decisão é tomada com recurso a um modelo de linguagem natural, e o resultado é depois tratado pela API, que comunica diretamente com o agente designado. Os agentes especializados funcionam de forma autónoma e recebem diretamente as perguntas a partir da API. A resposta gerada é então armazenada e devolvida ao utilizador.

Elementos inicialmente planeados, como o “MetaAgent”, “DataLake”, camadas intermediárias de orquestração e *dashboard* de logs, foram eliminados ou não representados, dado o seu impacto limitado na versão funcional do sistema.

Esta versão final assegura uma comunicação clara e eficiente, mantendo a escalabilidade do sistema e permitindo fácil manutenção e extensão futura com novos agentes ou funcionalidades.

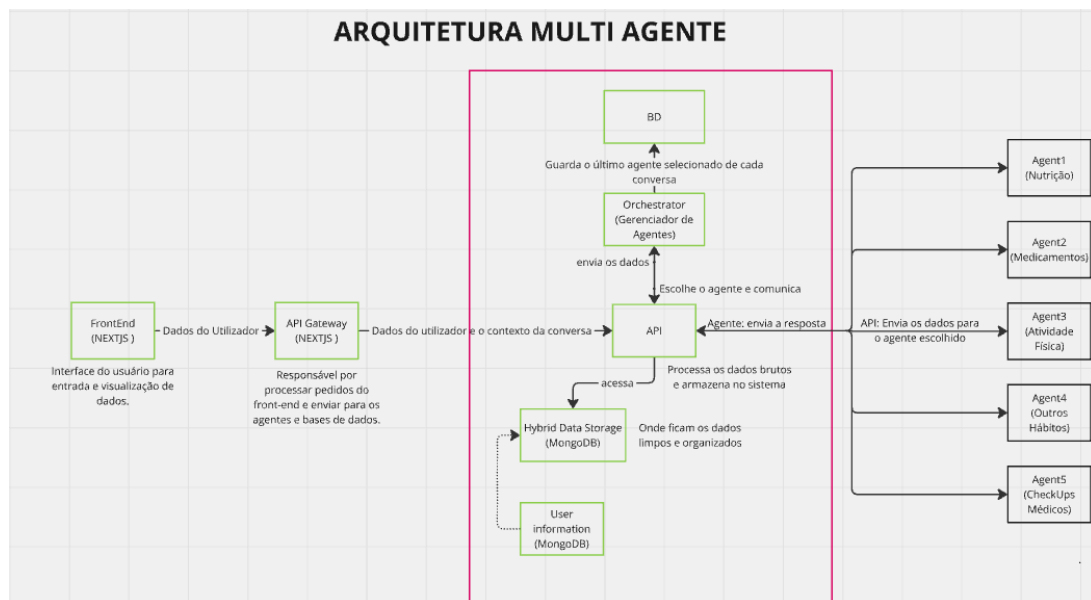


Figura 8: Arquitetura Final Produzida

4.2.1) Global Agent

O agente global (Global Agent) é uma parte fulcral do sistema, tendo como função associar a cada pergunta o agente especializado mais indicado para providenciar uma resposta. Assim, este agente entra em ação quando uma pergunta é colocada pelo utilizador. É importante destacar

que diferenças no funcionamento do agente ocorrem, durante uma conversa, quando recebe a primeira mensagem ou qualquer outra que não a primeira.

No primeiro caso, isto é, quando recebe a mensagem inicial de uma conversa, o agente global, mediante uma ligação API a uma LLM própria, determina qual o agente especializado correspondente, devolvendo apenas o nome do agente selecionado para a API. Nesta situação, o agente global tem, obrigatoriamente, que devolver o nome de um dos especialistas, mesmo que a pergunta não encaixe completamente em nenhum. Além disso, neste passo é gerada também uma *thumbnail*, ou seja, um pequeno resumo do tema da conversa (4-8 palavras) que será apresentado ao utilizador.

Já no caso oposto, ou seja, quando recebe uma mensagem de uma conversa já existente, o agente global também irá determinar qual o agente especializado mais indicado para responder à mensagem. Contudo, não será gerada uma nova *thumbnail* e, numa eventual situação em que reconheça que nenhum agente especializado seja adequado para responder à pergunta colocada, o agente global pode responder com “None”, informando a API que não foi encontrado um agente adequado e que deverá utilizar o agente correspondente à última pergunta colocada.

O agente global mantém, na sua base de dados, um registo de todas as conversas em que participa, de modo a verificar se a mensagem que recebe é a primeira de uma conversa ou faz parte de uma conversa já existente.

Todas as repostas providenciadas pelo agente global (escolha do agente e desenvolvimento da *thumbnail* da conversa) são obtidas através do modelo *Llama-3.3-70B-Instruct-Turbo*, disponibilizado através da API da *Together AI*.

4.2.2) Bases de Dados

A arquitetura do sistema recorre a uma abordagem baseada em bases de dados documentais, utilizando *MongoDB* como sistema principal de armazenamento. Para além da base de dados do Global Agent previamente mencionada, foram definidas duas componentes principais disponíveis à API: a base de dados do Orchestrator e o Hybrid Data Storage, cada uma com responsabilidades distintas e complementares.

A base de dados do Orchestrator é responsável por armazenar o último agente selecionado para cada conversa iniciada pelo utilizador, definido pelo agente global. Esta estratégia permite garantir consistência no fluxo das interações, assegurando que perguntas subsequentes dentro da mesma conversa são encaminhadas para o mesmo agente especializado, exceto se a situação justificar uma mudança. Este registo reduz a ambiguidade e reforça a coerência temática das respostas, simplificando a lógica da API e evitando reavaliações desnecessárias por parte do Global Agent.

Por outro lado, o Hybrid Data Storage representa a estrutura central onde se encontram armazenados a maioria dos dados do sistema. Esta base de dados também implementada com *MongoDB* contém, de forma estruturada, os dados do utilizador, incluindo perfis, preferências, condições clínicas, medicamentos, alergias e dietas, bem como o histórico completo das conversas e mensagens trocadas entre o utilizador e os agentes. A separação lógica entre perfis e interações é gerida por uma abstração que garante segurança, integridade e flexibilidade no acesso a estes dados.

A adoção deste modelo híbrido permite um balanço entre especialização e centralização, assegurando a rastreabilidade das interações e facilitando futuras extensões do sistema, nome-

adamente a inclusão de novos agentes ou funcionalidades personalizadas com base no perfil individual de cada utilizador.

4.2.3) Rotas e Fluxo

A API do projeto *NutriBot* foi desenvolvida em *Python*, recorrendo à *micro-framework Flask*, que permite a criação rápida e eficiente de aplicações *web*. Para facilitar a comunicação entre diferentes domínios, foi integrado o pacote *flask-cors*, que permite o acesso à API a partir de clientes externos, como aplicações web ou móveis. A gestão de variáveis de ambiente é feita através de *python-dotenv*, que carrega automaticamente as configurações sensíveis a partir de ficheiros *.env*, das quais fazem parte as informações relativas à base de dados documental, a porta em que a API vai estar disponível, a porta base de resposta dos agentes especialistas que será utilizada aquando dos pedidos de questões e a rota do Global Agent de forma a ser possível consultar eficientemente o mesmo.

A estrutura da API está organizada em *blueprints*, que separam as rotas relacionadas com perfis de utilizador (*profile*) das rotas de conversa (*chat*), promovendo assim uma arquitetura modular e de fácil manutenção. O acesso e manipulação dos dados são realizados através de uma ligação a uma base de dados *MongoDB*, encapsulada numa classe própria (*MongoDBConnection*), garantindo a abstração das operações de base de dados.

A API disponibiliza *endpoints RESTful* para criar, consultar, atualizar e eliminar perfis de utilizador, bem como para gerir conversas e mensagens. Todos os dados recebidos são validados quanto ao tipo e formato, assegurando a integridade da informação. O sistema suporta ainda enumerações para campos categóricos, listas para doenças, medicamentos, alergias e dietas, e validações específicas para cada campo do perfil.

Assim, a API do *NutriBot* utiliza tecnologias modernas e boas práticas de desenvolvimento em *Python*, com *Flask*, integração de *CORS*, gestão de variáveis de ambiente, ligação a *MongoDB* e uma arquitetura modular baseada em *blueprints*, proporcionando uma base robusta e escalável para aplicações deste contexto.

4.2.3.1) Fluxo de Questão do Utilizador

Este diagrama representa a sequência de interações, desde a criação de uma conversa até à resposta por parte do agente especializado, entre os componentes do sistema de *chat*:

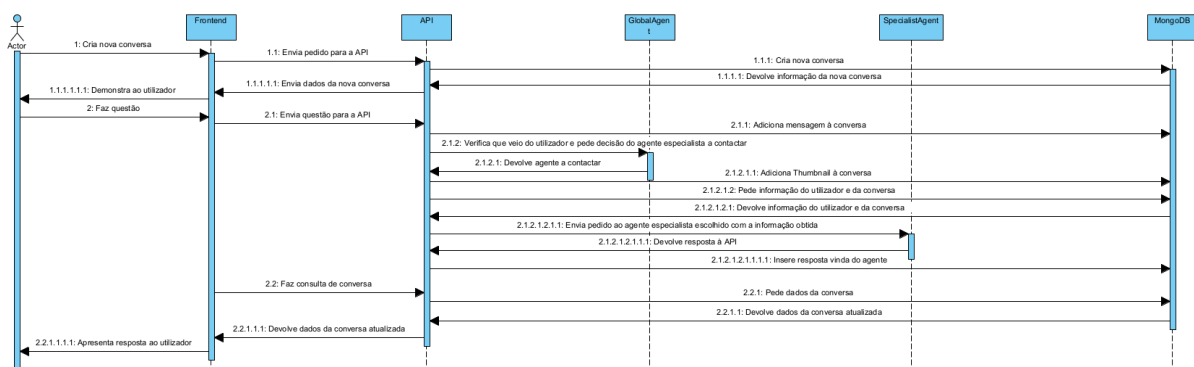


Figura 9: Diagrama de Sequência de Conversa

Para proceder à criação de uma nova conversa, o utilizador inicia um novo *chat*. O *frontend* envia o pedido à *API*, que por sua vez comunica com a classe que faz estabelecer comunicação

com o MongoDB para criar a conversa e armazená-la. O sistema devolve os dados da nova conversa ao **frontend**, para que os apresente ao utilizador.

Depois, o utilizador coloca uma questão. O **frontend** envia a questão para a **API**, que identifica o agente apropriado através do **GlobalAgent**. Sendo a primeira pergunta, adiciona a *thumbnail* gerada pelo **GlobalAgent** à conversa. A **API** recolhe os dados necessários, do utilizador e da conversa, e reencaminha tudo para o agente apropriado identificado, o **SpecialistAgent**, que posteriormente devolve uma resposta. Esta é guardada com o papel de *BOT* no MongoDB.

Para visualizar as respostas, o utilizador atualiza a página, o **frontend** pede à **API** os dados da conversa. A **API** consulta o MongoDB e devolve os dados atualizados ao **frontend**, que os mostra ao utilizador.

4.2.3.2) Fluxo de Configuração de Dados do Perfil do Utilizador

O diagrama de sequência seguinte descreve o processo de visualização e edição dos dados do perfil de um utilizador numa aplicação. O processo envolve quatro participantes principais: o **Actor** (utilizador), o **Frontend**, a **API** e a base de dados **MongoDB**.

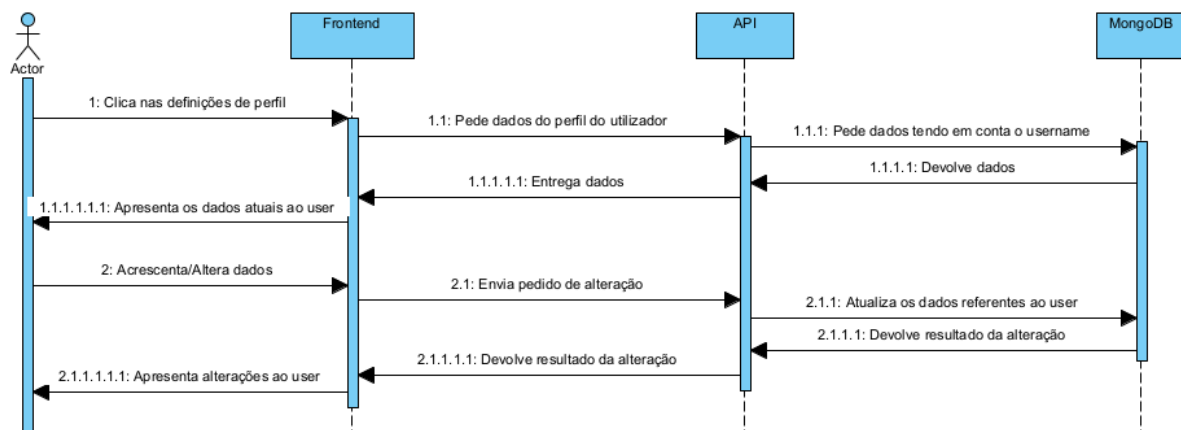


Figura 10: Diagrama de Sequência de Configuração de User

O processo começa quando o utilizador clica nas definições de perfil, o que leva o **Frontend** a solicitar os dados do perfil do utilizador. Este pedido é então encaminhado para a **API**, que, com base no *username* do utilizador, requisita os dados necessários à base de dados **MongoDB**.

Uma vez obtidos, os dados são devolvidos pela base de dados à **API**, que os transmite ao **Frontend**. De seguida, o **Frontend** apresenta as informações atuais ao utilizador.

Com os dados já visíveis, o utilizador pode optar por acrescentar ou alterar informações do seu perfil. Essas alterações são enviadas do **Frontend** para a **API**, que procede à atualização dos dados na base de dados **MongoDB**.

Após a atualização ser realizada, a base de dados envia o resultado da operação de volta à **API**, que o reencaminha para o **Frontend**. Finalmente, o **Frontend** apresenta ao utilizador o resultado da alteração, refletindo as novas informações introduzidas.

5) Role C

A presente secção documenta o desenvolvimento da componente Role C do projeto, que consiste na criação de uma base de conhecimento científica estruturada e robusta. Esta componente visa sistematizar e disponibilizar informação rigorosa e atualizada nos domínios da nutrição, suplementação, atividade física, estilos de vida e medicina preventiva. A base de conhecimento foi concebida para apoiar agentes especializados na produção de respostas contextualizadas, consistentes e fundamentadas em evidência científica. A sua arquitetura permite a integração de múltiplas fontes de informação e a operação coordenada de diferentes agentes especializados.

5.1) Arquitetura Geral do Sistema

A arquitetura da base de conhecimento foi estruturada como uma pipeline integrada de aquisição, processamento e gestão de informação científica. Esta abordagem sistemática permite o tratamento eficiente de grandes volumes de dados, desde a recolha automatizada em fontes credíveis até à transformação e organização em estruturas otimizadas para recuperação de informação.

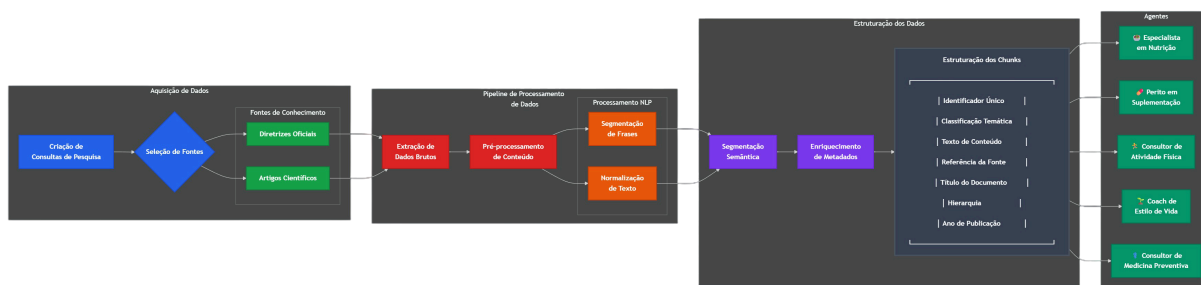


Figura 11: Arquitetura Geral do Sistema da Base de Conhecimento

O sistema é composto por três componentes fundamentais que operam de forma sequencial e coordenada¹:

- **Aquisição de Dados:** Responsável pela recolha de informação científica através de múltiplas metodologias, incluindo APIs especializadas, técnicas de webscraping, transferência de documentos PDF e outros métodos manuais ou semi-automatizados, utilizando queries temáticas direcionadas para maximizar a relevância dos dados obtidos.
- **Pipeline de Processamento de Dados:** Aplica processos de limpeza, normalização e validação dos dados brutos, assegurando a qualidade e consistência da informação antes da estruturação.
- **Estruturação dos Dados:** Realiza a segmentação semântica dos dados processados e enriquece a informação com metadados estruturados, preparando-a para posterior carregamento. Esta fase inclui a implementação de sistemas de classificação e priorização de fontes, permitindo a atribuição de pesos diferenciados com base na credibilidade e relevância científica.

¹O carregamento dos dados estruturados na base de conhecimento constituiu uma fase adicional cuja responsabilidade coube a cada grupo de trabalho definir, podendo ser realizada pela Role C ou pela role responsável pela implementação do sistema RAG, conforme as necessidades específicas de cada domínio temático.

5.2) Metodologia de Extração de Dados

O processo de extração inicia-se com a definição de queries específicas para cada domínio temático, desenvolvidas pelos respetivos grupos de trabalho. Esta abordagem dirigida permite otimizar a recolha de literatura científica relevante, minimizando a inclusão de informação irrelevante ou de baixa qualidade.

- **Nutrição - Hábitos alimentares**
"Eating habits" AND "nutrition" AND ("health outcomes" OR "dietary patterns")
- **Suplementos alimentares e outros fármacos para prevenção de doenças**
"preventive effects of dietary supplements on chronic diseases."
- **Atividade física - caminhada, exercício físico, ginásio, desporto, etc**
"physical activity and health benefits"
- **Outros hábitos: sono, horas de trabalho/descanso, ergonomia, higiene, tabaco, etc**
"sleep quality and health outcomes"
- **Check-ups/acompanhamento médico de medicina preventiva**
"preventive medical check-ups effectiveness"

Figura 12: Exemplos de Queries Formuladas por Domínio Temático

A extração é realizada através de múltiplas metodologias, incluindo integração com APIs, técnicas de webscraping e processos manuais de transferência de documentos. As principais fontes incluem:

- **PubMed:** Base de dados biomédica da National Library of Medicine, fornecendo acesso a literatura médica peer-reviewed
- **Google Scholar:** Plataforma de pesquisa académica que abrange múltiplas disciplinas e tipos de publicações
- **Repositórios Indexados:** Bases de dados especializadas em nutrição, medicina preventiva e ciências da saúde

A seleção das fontes obedece a critérios rigorosos de credibilidade, atualidade, abrangência e relevância científica, assegurando a qualidade da informação integrada na base de conhecimento.

5.3) Processamento e Transformação da Informação

O processamento dos documentos científicos obtidos segue uma metodologia sistemática de transformação que garante a qualidade, coerência e normalização da informação. Este processo compreende várias etapas sequenciais:

- **Limpeza Textual:** Remoção de elementos que não contribuem para o conteúdo informativo, incluindo caracteres especiais inválidos, formatação inconsistente e elementos de markup irrelevantes.
- **Normalização Linguística:** Uniformização de convenções textuais, incluindo acentuação, pontuação, espaçamento e capitalização, de forma a garantir consistência na representação textual.
- **Segmentação Inteligente:** Divisão do conteúdo em unidades temáticas coerentes (chunks), preservando a integridade semântica e facilitando a recuperação posterior de informação específica.

Esta fase visa não apenas a padronização dos dados, mas também a preparação otimizada para mecanismos subsequentes de indexação, classificação e inferência. A transformação assegura que a informação seja armazenada numa forma que maximize a eficiência dos processos de recuperação e geração de respostas.

5.4) Estruturação dos Dados

Cada unidade textual (chunk) é enriquecida com metadados estruturados que facilitam a recuperação e contextualização da informação:

<u>chunk_id</u>	Identificador único para cada chunk .
<u>chunk_text</u>	Conteúdo textual em si.
<u>title</u>	Título da seção ou do documento de onde o chunk foi extraído.
<u>link</u>	URL ou referência da fonte original.
<u>year</u>	Ano de publicação ou atualização da fonte.
<u>topic</u>	Tópico principal associado ao chunk .
<u>hierarchy</u>	Indica a credibilidade da fonte.

Figura 13: Estrutura de Metadados Associados a Cada Chunk

Durante a fase de estruturação, é aplicado um sistema de hierarquização de fontes que classifica a informação em três níveis de credibilidade científica:

- **Nível 1 - Diretrizes Oficiais:** Documentos e recomendações emitidas por entidades oficiais reconhecidas nacional e internacionalmente, tais como a Organização Mundial de Saúde (OMS), Direção-Geral da Saúde (DGS) e outras autoridades de saúde pública.
- **Nível 2 - Literatura Científica Peer-Reviewed:** Artigos científicos publicados em revistas indexadas e sujeitas a revisão por pares, sob a forma de abstracts.
- **Nível 3 - Documentação Técnica Complementar:** Relatórios de investigação, documentos técnicos e outras fontes de informação científica credível.

Esta classificação hierárquica é integrada nos metadados de cada chunk através do campo *hierarchy*, permitindo a atribuição automática de pesos diferenciados durante os processos de recuperação de informação e priorização de conteúdos.

A validação final é realizada através de uma abordagem distribuída e colaborativa, onde os grupos de trabalho temáticos asseguram a integridade dos dados e a completude dos metadados. Este processo garante que apenas informação verificada e contextualizada seja integrada na base de conhecimento.

5.5) Conclusões e Trabalho Futuro

O desenvolvimento da base de conhecimento externo estabeleceu uma infraestrutura robusta para a gestão de informação científica especializada, apresentando três contribuições fundamentais: um modelo de hierarquização de fontes que assegura a credibilidade da informação através de evidência científica de qualidade; uma pipeline de processamento otimizada para segmentação e organização sistemática de dados compatível com tecnologias RAG; e uma estrutura unificada de dados, adaptável e escalável, que facilita a integração de múltiplos domínios temáticos e a operação coordenada de agentes especializados.

A análise crítica do trabalho revela limitações que constituem oportunidades de desenvolvimento. A ausência de validação empírica com utilizadores reais e a falta de benchmarking comparativo com soluções existentes impedem uma avaliação completa da eficácia e performance do sistema em contextos práticos.

As perspetivas futuras contemplam a implementação de algoritmos de ranking inteligentes baseados em machine learning para priorização sofisticada da informação e o desenvolvimento de sistemas de personalização dinâmica adaptados a perfis de utilizador. Estas direções de investigação visam consolidar a plataforma como referência na gestão de conhecimento científico especializado, expandindo significativamente as suas capacidades operacionais e impacto prático.

6) Role D

A Role D foi responsável pela definição e implementação da estrutura de dados dos utilizadores da aplicação desenvolvida, assegurando que toda a informação necessária ao funcionamento dos agentes especialistas estivesse disponível de forma organizada.

Para isso, desenvolvemos um modelo utilizando *MongoDB*. O paradigma documental foi escolhido principalmente pela sua versatilidade e flexibilidade em relação ao *schema*. Além de escalar horizontalmente, também torna a base de dados facilmente extensível. Assim, a qualquer momento é possível se necessário adicionar mais informação ao sistema.

Passamos a descrever as entidades desenvolvidas, acompanhadas com o seu propósito. Foi também desenvolvido um diagrama de classes para uma mais fácil compreensão desta proposta.

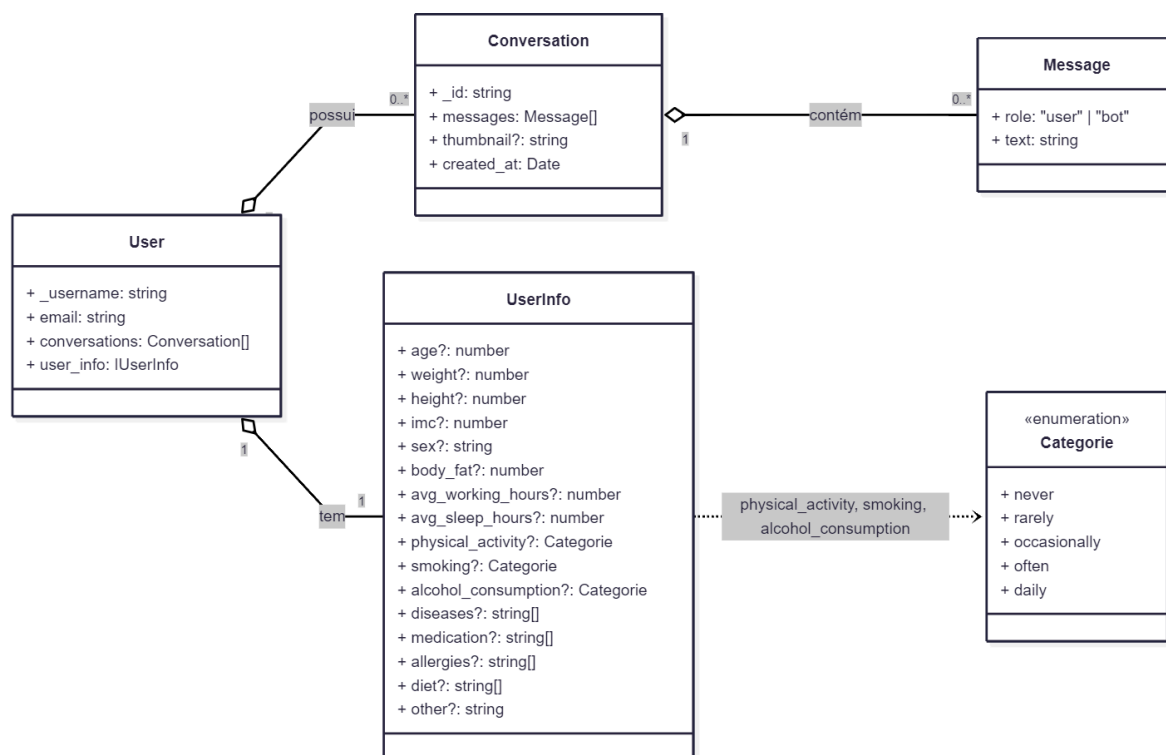


Figura 14: Diagrama de Classes do modelo de dados dos utilizadores

6.1) Entidades

6.1.1) User

- `__username` -> Campo **único** e **obrigatório** do tipo *String*
- `email` -> Campo **único** e **obrigatório** do tipo *String*
- `conversations` -> Uma lista do tipo *Conversation*
- `user_info` -> Estrutura do tipo *UserInfo*

Nesta entidade, os campos `username` e `email` têm o propósito de atuar como credenciais na aplicação final. Já `conversations` representa todas as conversas do utilizador e `user_info` a informação pessoal do utilizador, sendo este mais detalhado posteriormente.

6.1.2) Conversation

- `__id` -> Campo do tipo *ObjectId*
- `messages` -> Uma lista do tipo *Message*

- **thumbnail** -> Um campo do tipo *String*
- **created_at** -> Campo do tipo *Date*

Aqui, **messages** é referente a todas as mensagens de uma determinada conversa. A **thumbnail** serve para o utilizador ter um indicador sobre qual conversa se trata, sendo esta uma frase/resumo daquilo que é a conversa em questão. E finalmente, **created_at** sendo a data de início da conversa

6.1.3) Message

- **role** -> Campo **obrigatório** do tipo *String* (“**user**” | “**bot**”)
- **text** -> Campo **obrigatório** do tipo *String*

Para esta entidade, **role** só pode tomar o valor “**user**” ou “**bot**” e serve para identificar o emissor da mensagem e **text** referente ao conteúdo da própria mensagem

6.1.4) UserInfo

Para esta entidade, todos os campos são opcionais. Apesar da informação ser muitas vezes necessária e ajudar muitos nos resultados produzidos, além da questão de personalização da resposta, esta decisão transporta para o utilizador final uma maior confiança, possuindo um maior controlo dos seus dados, podendo sempre que quiser adicionar ou remover informação sobre si mesmo.

Além disto para a construção desta mesma entidade, foi necessário suportar as necessidades de todos os grupos. Para isso, cada elemento da **role** identificou junto do seu grupo a informação necessária para o seu tema. As informações para cada grupo vão passar a ser mais detalhadas.

- Grupo 1 - Nutrição:
 - Alergias
 - Tipo de Dieta
- Grupo 2 - Suplementos alimentares/fármacos para prevenção de doenças:
 - IMC
 - %Gordura Corporal
 - Atividade Física
 - Medicação
 - Alergias
- Grupo 3 - Atividade física:
 - IMC
 - Média de Horas de Sono
 - Consumo de Tabaco
 - Média de horas de trabalho
- Grupo 4 - Outros Hábitos:
 - Consumo de Tabaco
 - Média de Horas de Trabalho
 - Média de horas de Sono
- Grupo 5 - Acompanhamento Médico:
 - Doenças Crónicas
 - Medicação atual

- Alergias
- Consumo de álcool
- Atividade Física
- Além da informação detalhada acima para cada grupo, notou-se que havia alguma informação que era comum às necessidades de todos os grupos sendo essas:
 - Idade
 - Peso
 - Altura
 - Sexo

Este exercício além de nos permitir chegar a uma estrutura final consistente e completa, permite ainda que seja passada apenas a informação necessária para o agente que vai tratar da questão, não sobrecarregando com informação desnecessária sobre o utilizador.

- *age* -> Campo do tipo *Number*
- *weight* -> Campo do tipo *Number*
- *height* -> Campo do tipo *Number*
- *imc* -> Campo do tipo *Number*
- *sex* -> Campo do tipo *String*
- *body_fat*(%) -> Campo do tipo *Number*
- *avg_working_hours* -> Campo do tipo *Number*
- *avg_sleep_hours* -> Campo do tipo *Number*
- *physical_activity* -> Campo do tipo *Categorie*
- *smoking* -> Campo do tipo *Categorie*
- *alcohol_consumption* -> Campo do tipo *Categorie*
- *diseases* -> Uma lista do tipo *String*
- *medication* -> Uma lista do tipo *String*
- *allergies* -> Uma lista do tipo *String*
- *diet* -> Uma lista do tipo *String*
- *other* -> Campo do tipo *String*

Os campos do tipo *Categorie* como *physical_activity*, *smoking* e *alcohol_consumption* foram desenvolvidos como campos categóricos, isto é, só podem tomar um dos seguintes valores:

- “*never*” -> nunca
- “*rarely*” -> uma vez por mês ou menos
- “*occasionally*” -> duas a quatro vezes por mês
- “*often*” -> duas a três por semana
- “*daily*” -> diariamente

Esta foi uma forma de padronizar as respostas dos utilizadores quanto a estas informações, criando o máximo de objetividade possível, facilitando assim o tratamento e análise por parte dos agentes especialistas.

Neste caso, os campos *diseases*, *medication*, *allergies* e *diet* também têm um tratamento especial. Apesar de serem todas listas de *String*, o utilizador não pode colocar o que quiser nessa lista. Tal como no caso anterior, foi criada uma padronização. A solução passou por criar uma lista para cada campo, podendo o utilizador posteriormente escolher entre as opções que se encontram em cada lista. Estas listas são estáticas mas podem ser modificadas “*hardcoded*”

no caso de ser necessário atualizar as mesmas. Isto permite que não haja redundâncias nem valores imprevisíveis.

O único campo onde não há qualquer controlo é no ***other***, onde o utilizador pode colocar qualquer tipo de informação pessoal que achar relevante para o *bot*.

6.2) Limitações

A principal limitação passa pela imprevisibilidade do tamanho da base de dados. Isto porque, no caso de um recém utilizador, que tenha apenas criado conta, o espaço utilizado relativo ao seu perfil é quase que residual. Por outro lado, um utilizador que utilize muito intensivamente a aplicação vai naturalmente ocupar um espaço proporcional ao seu uso, já que a entidade relativa às conversas ocupará mais espaço. Com isto, é necessário implementar mecanismos ou utilizar serviços que permitam um grande dinamismo quanto ao tamanho da base de dados para que não haja problemas em relação ao espaço.

7) Role E

O Role E é responsável por gerir agentes que utilizam a abordagem de RAG (*Retrieval-Augmented Generation*) para lidar com tópicos específicos. Ele desempenha um papel crucial na organização e processamento de informações, garantindo que os dados relevantes sejam integrados de forma eficiente para gerar respostas precisas.

7.1) Workflow

Para cumprir esse objetivo, o agente segue um *workflow* estruturado que será descrito a seguir. Este *workflow* organiza o fluxo de informações desde o recebimento da solicitação inicial até a entrega da resposta final. Nas próximas seções, cada etapa do processo será detalhada.

O processo começa com o recebimento do *payload* inicial, que contém as informações básicas da solicitação. Em seguida, ocorre a agregação de dados utilizando a técnica de RAG (*Retrieval-Augmented Generation*), que permite enriquecer o contexto da solicitação com informações relevantes. Após isso, é realizada a geração do *prompt* final, que será enviado ao modelo de linguagem. Na etapa seguinte, a pergunta é enviada ao LLM (*Large Language Model*), que processa o *prompt* e gera uma resposta. Por fim, a resposta é retornada ao utilizador, completando o ciclo do *workflow*. De seguida iremos abordar cada fase em detalhe.



Figura 15: Workflow do Agente Especialista

7.1.1) Payload Inicial

A etapa inicial consiste na receção do *payload*, que contém as informações essenciais para o processamento da solicitação. Este *payload* inclui o *username* e o id da conversa, que identificam unicamente o pedido. Além disso, são fornecidas informações do utilizador, como o histórico de conversas anteriores e dados pessoais relevantes, que ajudam a contextualizar a solicitação. Por fim, o *payload* também contém a pergunta a ser efetuada, que será processada nas etapas seguintes.

7.1.2) RAG (Inserção e Vetorização)

A etapa de inserção e vetorização no processo de RAG é fundamental para preparar os dados que serão utilizados na geração de respostas. Para todo o processo de guardar e consultar dados foi utilizado o **Pinecone**. Esta escolha deve-se à sua eficiência na gestão de vetores de alta dimensão, oferecendo uma solução escalável e de baixa latência para armazenamento e recuperação de *embeddings*. Este processo é composto por três fases principais:

1. **Leitura e preparação de dados:** Os dados são lidos a partir de arquivos no formato JSON, que contêm informações estruturadas, como texto, título, links e outros metadados relevantes.
2. **Criação de *embeddings*:** Nesta fase, os textos são transformados em representações vetoriais (*embeddings*) utilizando um modelo especializado, neste caso o `llama-text-embed-v2`. O tipo de *input* utilizado é do tipo **passage**, ou seja, trechos de texto que serão convertidos em vetores numéricos. Estes vetores capturam as características semânticas dos textos, permitindo que sejam comparados de forma eficiente e eficaz.
3. **Criação dos vetores para indexação:** Após os *embeddings* serem gerados, os vetores são organizados para indexação. Cada vetor tem a si associado a um identificador único (id), que representa um *chunk* específico do texto. Além disso, são armazenados os valores do vetor de *embedding* e metadados adicionais, como texto, título, ano e tópico. Estes elementos são essenciais para facilitar a recuperação de informações relevantes durante o processo de consulta.

7.1.3) RAG (Recolha de Dados)

O processo inicia-se com a geração do *embedding* da pergunta, que é utilizado para realizar uma pesquisa hierárquica. A pesquisa começa no nível mais alto da hierarquia, onde são recolhidos os N melhores resultados.

Os resultados obtidos são avaliados com base no seu *score*. Se o *score* de um resultado for maior ou igual ao *score* ideal, ele é adicionado à lista de *chunks* finais. Caso contrário, o algoritmo verifica se existe um nível inferior na hierarquia ao qual pode consultar mais informação. Este ciclo repete-se até que não existam mais níveis inferiores ou até que os *chunks* finais atinjam o número desejado (*topK chunks*).

Quando os *chunks* finais são determinados, o algoritmo aplica um filtro adicional para garantir que todos os resultados tenham um **score** maior ou igual a um *score* mínimo. Por fim, os dados extraídos são utilizados para construir a resposta, combinando informações como título, ano, link e texto.

A imagem abaixo apresenta o fluxo detalhado do processo de recolha de dados no RAG, destacando as etapas principais e as condições avaliadas em cada fase.

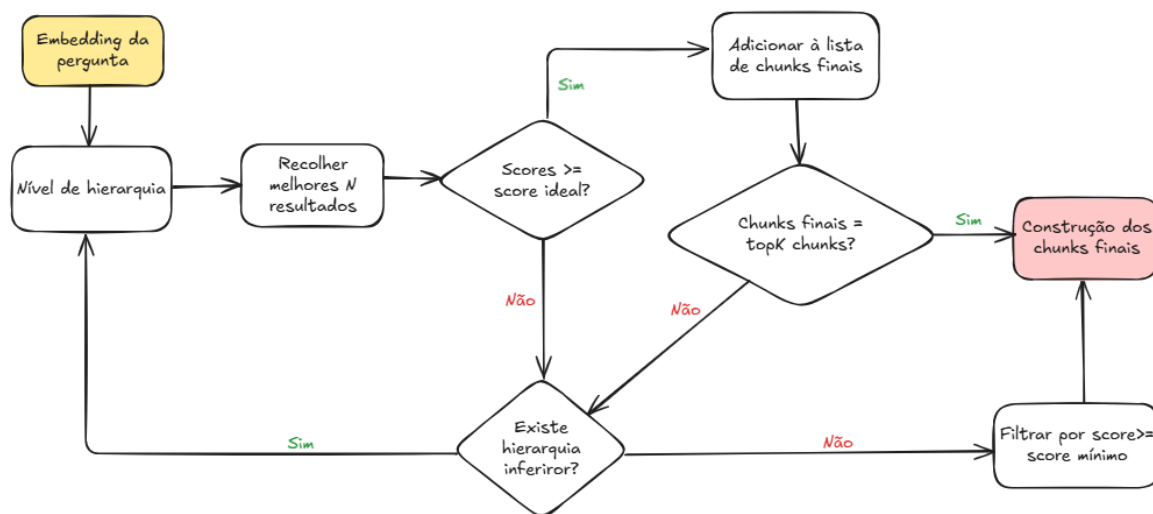


Figura 16: Processo de Recolha de Dados no RAG

7.1.4) Gerar *Prompt* Final

A etapa de gerar o *prompt* final é crucial para consolidar todas as informações relevantes que serão enviadas para o LLM. Este *prompt* é composto por diversos elementos que garantem que a resposta gerada seja precisa e contextualizada.

Em primeiro lugar, são incluídas as regras de conduta, que definem os limites e orientações para o comportamento do modelo, assegurando que as respostas estejam alinhadas com os objetivos do sistema. Em seguida, a pergunta do utilizador é incorporada, representando o núcleo da solicitação que será processada.

Além disso, são adicionadas informações pessoais do utilizador, como dados relevantes para personalizar a resposta, e o histórico de conversa (se possível), que fornece o contexto necessário para compreender a interação em andamento. Por fim, o contexto é enriquecido com os *chunks* selecionados na etapa anterior, que contêm as informações mais relevantes para fundamentar a resposta.

O algoritmo responsável pela geração do *prompt* final também considera o número máximo de *tokens* permitidos pelo modelo de linguagem. Este limite é essencial para garantir que o *prompt* seja processado corretamente, respeitando as restrições do modelo. Caso o histórico de conversa ou o contexto selecionado sejam muito extensos, o algoritmo aplica uma estratégia de redução progressiva.

Inicialmente, tenta-se incluir o máximo de informações possíveis, priorizando os elementos mais relevantes, como a pergunta do utilizador e os *chunks* selecionados. Se o número de *tokens* exceder o limite permitido, o algoritmo começa a reduzir o histórico de conversa, removendo gradualmente as interações mais antigas. Este processo continua até que o *prompt* final esteja dentro do intervalo válido de *tokens*, garantindo que o modelo receba um *prompt* otimizado e que preserve as informações mais relevantes para a geração de uma resposta de qualidade.

7.1.5) Enviar Pergunta para o LLM

Nesta etapa, a pergunta gerada no *prompt* final é enviada para um LLM para processamento. Existem dois modelos disponíveis para este processo:

- **Llama-3.3-70B**: Um modelo robusto e eficiente para diversas tarefas de linguagem natural.
- **Deepseek R1 70B (Reasoning)**: Um modelo especializado em tarefas de raciocínio, ideal para cenários que exigem maior capacidade analítica.

Para a integração com os modelos, utilizamos a plataforma **Together.ai**, que oferece acesso gratuito aos modelos. No entanto, esta solução possui uma limitação de 5 requisições por minuto.

7.1.6) Retornar Resposta

A etapa final do processo consiste em retornar a resposta ao utilizador. Esta resposta contém informações essenciais para identificar a interação e comunicar o resultado do processamento. Os elementos incluídos são:

- **Conversation Id**: Identificador único da conversa, que permite rastrear a interação.
- **Username**: Nome do utilizador que realizou a solicitação.
- **Mensagem gerada**: Em caso de sucesso, a resposta gerada pelo modelo é enviada.
- **Erro**: Em caso de falha, é retornada uma mensagem de erro detalhada, indicando o motivo do problema.

Este formato garante que o utilizador receba um feedback claro e estruturado, independentemente do resultado do pedido.

7.2) Configuração

Cada grupo define as especificações do seu agente num ficheiro JSON, ajustando os parâmetros de acordo com as necessidades específicas do seu caso de uso. Estes parâmetros permitem personalizar o comportamento do agente e otimizar o processamento das solicitações. Os campos disponíveis incluem:

- **Reasoning Model**: Indica se o modelo de *reasoning* será utilizado ou não.
- **TopK chunks**: Define o número máximo de *chunks* a serem considerados.
- **Min Threshold**: Especifica o *score* mínimo aceitável para os resultados.
- **Target Threshold**: Determina o *score* ideal para a seleção de resultados.
- **Max Hierarchical Level**: Limita o número de níveis hierárquicos a serem explorados.
- **Context Prompt**: Caminho para o ficheiro que contém o *prompt* de contexto.
- **Chunked Data**: Caminho para o ficheiro que contém os dados segmentados (*chunks*).

7.3) Benchmarking

O processo de benchmarking foi realizado para avaliar a qualidade das respostas geradas por diferentes modelos de linguagem. Testámos vários modelos, incluindo **ChatGPT-4.0**, **Claude Sonnet 4** e **Gemini 2.5 Pro**, comparando o desempenho entre configurações com raciocínio (*reasoning*) e sem raciocínio.

Cada modelo foi avaliado em diferentes cenários e recebeu notas de 1 a 10, com base na relevância e precisão das respostas. Este processo permitiu identificar as forças e fraquezas de cada modelo em contextos específicos.

Os resultados obtidos foram organizados e serão partilhados mais à frente, na secção dedicada a cada grupo, visto que cada grupo possui os seus próprios resultados. Esta abordagem permite

uma análise detalhada e específica, facilitando a comparação entre os diferentes agentes e as suas configurações.

8) Especialista em Nutrição

8.1) Análise do tema e das fontes de dados relevantes

A nutrição e os hábitos alimentares são temas centrais para a saúde pública, especialmente diante do aumento das doenças crônicas relacionadas à má alimentação, como obesidade, diabetes tipo dois e problemas cardiovasculares. A análise dos padrões alimentares da população permite identificar não apenas os alimentos consumidos, mas também os comportamentos e fatores que influenciam estas escolhas, como disponibilidade, cultura, e conhecimento nutricional. Estudar os hábitos alimentares possibilita o desenvolvimento de estratégias eficazes para promover mudanças comportamentais, aumentar a conscientização e melhorar a qualidade de vida, refletindo diretamente na prevenção de doenças.

Para além das fontes de dados referidas anteriormente pela Role C, foram recolhidas informações de fontes mais específicas ao tema. Um exemplo foi o *EatRight* (eatright.org) site oficial da *Academy of Nutrition and Dietetics*. Este site tem por objetivo promover a saúde por meio de informações baseadas em evidências sobre alimentação, nutrição e bem-estar. É uma fonte confiável de informação nutricional baseada na ciência, útil tanto para quem quer aprender a comer melhor quanto para profissionais que atuam com saúde, nutrição e dietética. Assim, achamos que poderia ser uma fonte útil. Outro exemplo é o *Dietary Guidelines for Americans* (dietaryguidelines.gov) que representam as diretrizes alimentares oficiais dos Estados Unidos, desenvolvidas conjuntamente pelos Departamentos de Agricultura (USDA) e Saúde e Serviços Humanos (HHS). Estas *guidelines*, constituem a base científica para políticas nutricionais federais, programas de alimentação escolar, e recomendações clínicas. A sua importância reside no facto de representarem o consenso científico mais atual sobre padrões alimentares saudáveis, baseando-se numa revisão exaustiva da literatura científica disponível e em análises de comités de especialistas independentes. A integração desta fonte no sistema permite acesso direto às recomendações oficiais que influenciam práticas nutricionais a nível populacional.

8.2) Metodologia de Recolha e Processamento de Dados

8.2.1) Estratégia de Web Scraping

A recolha automatizada de dados das fontes identificadas requer uma abordagem técnica robusta que combine diferentes ferramentas e metodologias. O processo foi estruturado em várias fases para garantir a qualidade e consistência dos dados extraídos.

8.2.2) Ferramentas de Extração

O Selenium WebDriver foi escolhido como ferramenta principal para a extração de dados, especialmente para fontes que requerem interação dinâmica, como é comum em plataformas académicas modernas. Esta escolha justifica-se pela capacidade de simular comportamento humano real, contornando potenciais bloqueios anti-bot e permitindo navegação através de interfaces complexas. Para fontes mais simples ou APIs disponíveis, foram também utilizadas bibliotecas *Python* especializadas:

- BioPython (Entrez): Interface oficial para acesso ao PubMed via NCBI,
- Requests: Para requisições HTTP diretas às APIs disponíveis,
- PyMuPDF (fitz): Extração de texto de documentos PDF,
- python-docx: Processamento de documentos Microsoft Word,
- Rich Console: Interface de utilizador aprimorada para monitorização do processo, etc.

Para o processamento de Linguagem Natural utilizamos:

- spaCy: Pipeline completo de NLP com modelo transformer pré-treinado,
- SentenceTransformers: Geração de embeddings semânticos,
- BAAI/bge-small-en-v1.5: Modelo específico para representação textual

8.2.3) Utilização de queries

Para a extração dos dados definimos algumas queries, seguem-se alguns exemplos:

```
"eating habits" AND "nutrition" AND ("health outcomes" OR "dietary patterns")
```

```
"dietary intake" AND "eating habits" AND "nutrition"
```

```
"plant-based diet" AND "eating behavior" AND ("health benefits" OR "disease prevention")
```

```
"nutrition education" AND "eating habits" AND ("prevention" OR "public health")
```

8.2.4) Definição de Keywords por Domínio

A estratégia de busca baseia-se também em taxonomias cuidadosamente estruturadas:

- Dietas e Padrões Alimentares,
- Hábitos e Comportamentos Alimentares,
- Nutrientes e Componentes Alimentares,
- Alimentos Específicos

De seguida, seguiu-se o Processamento, Chunking Semântico, Armazenamento dos dados em Pinecone (e ficheiros JSON) e a Indexação, que por serem semelhantes em todos os grupos foram explorados nos tópicos anteriores (role C) do relatório.

8.3) Avaliação do pipeline RAG

8.3.1) Objetivos do Teste

O teste teve como objetivo avaliar a performance do sistema RAG na recuperação de informação e geração de resposta factual com base em artigos fornecidos. Especificamente, pretendeu-se verificar a capacidade do sistema em:

- Identificar e recuperar documentos relevantes para responder a questões específicas sobre nutrição,
- Gerar respostas precisas e bem fundamentadas utilizando exclusivamente o conteúdo recuperado,
- Manter coerência e alinhamento com as instruções de *prompting* fornecidas,
- Citar adequadamente as fontes utilizadas na construção da resposta.

8.3.2) Questão e Resultado Obtido

Questão testada: “*What type of fat is avocado rich in?*”

Resposta gerada pelo sistema: “*Avocados are rich in monounsaturated fat, a type of unsaturated fat that is considered heart-healthy. This information is supported by multiple*

sources, including articles from the Academy of Nutrition and Dietetics and the World Health Organization, which emphasize the benefits of monounsaturated fats for cholesterol levels and overall cardiovascular health.”

A resposta demonstrou precisão factual, clareza na comunicação e adequado suporte bibliográfico, respondendo diretamente à questão formulada.

8.3.3) Recuperação

O sistema efetuou uma busca hierárquica até ao nível 3 de profundidade, recuperando 10 *chunks* relevantes com *scores* de similaridade variando entre 0.5233 e 0.3799. Apesar dos *scores* de similaridade relativamente baixos (máximo 0.52), o mecanismo de recuperação identificou corretamente documentos contendo informação diretamente relacionada com a questão.

Os documentos recuperados incluíram múltiplas referências ao conteúdo de gorduras monoin-saturadas em abacates, provenientes de fontes credíveis como:

- Academy of Nutrition and Dietetics (EatRight)
- World Health Organization (WHO)
- Artigos especializados em saúde cardiovascular

O *chunk* com maior score (ce7b7ade-9dda-4efe-9c10-34cd69a8f39a_chunk_6, Score: 0.5233) continha a informação mais direta: “*Avocados not only contain monounsaturated fat, but also are packed with dietary fiber, potassium and vitamins.*”

8.3.4) Geração

O LLM gerou uma resposta correta, clara e alinhada com as instruções de *prompting*. A resposta baseou-se exclusivamente no conteúdo fornecido, sem recorrer a conhecimento externo. O modelo demonstrou capacidade de:

- Sintetizar informação de múltiplos *chunks* de forma coerente,
- Identificar o tipo específico de gordura (monoinsaturada) presente no abacate,
- Contextualizar a informação dentro do âmbito da saúde cardiovascular,
- Manter rigor científico na formulação da resposta,
- Citar adequadamente as organizações de referência (Academy of Nutrition and Dietetics e WHO).

8.3.5) Observações Técnicas

Os scores de similaridade baixos podem indicar oportunidades de melhoria ao nível dos *embed-dings*, *chunking* ou *thresholds* de recuperação. No entanto, o sistema demonstrou robustez ao conseguir extrair informação correta mesmo sob essas condições, evidenciando a eficácia do processo de ranking e seleção de documentos.

A diversidade de fontes recuperadas (EatRight, WHO) fortaleceu a credibilidade da resposta final, permitindo ao sistema construir uma resposta bem fundamentada e cientificamente sólida. O facto de diferentes *chunks* complementarem a informação principal demonstra a eficácia da estratégia de recuperação hierárquica implementada.

8.3.6) Avaliação Quantitativa

Fase	Avaliação	Comentário
Recuperação (retrieval)	7/10	Recuperou os documentos corretos, mas os scores baixos indicam espaço para otimização de embeddings e chunking.
Processamento de contexto (RAG assembly)	8/10	Montou um contexto adequado e relevante, com boa filtragem de documentos não relevantes.
Geração de resposta (LLM)	10/10	Resposta correta, factual, alinhada com o <i>prompt</i> , sem alucinações, cumpre as regras.
Score Global	8.5/10	Sistema funcional e robusto, com boa resiliência a pequenas imperfeições na recuperação.

8.4) *Benchmarks* de Modelos

Nesta secção, apresenta-se a fundamentação da escolha do modelo adotado para a geração de respostas, tendo em conta duas variantes disponíveis: um modelo com capacidade de raciocínio (Reasoning Model) e outro sem essa componente (Non Reasoning Model).

Para tomar esta decisão, como já mencionado, foi realizada uma avaliação comparativa entre ambos os modelos, recorrendo à análise da qualidade das respostas geradas. Essa avaliação foi feita com o apoio de modelos de linguagem externos, *ChatGPT-4.0*, *Claude Sonnet 4*, *Gemini 2.5 Flash* e *Grok 3*, que classificaram as respostas de cada modelo numa escala de 1 a 10, com base em critérios como clareza, precisão, completude e relevância da informação.

8.4.1) Avaliação da Qualidade de Resposta dos Modelos

Para avaliar o desempenho dos modelos, foram elaboradas perguntas distribuídas por diferentes tópicos temáticos relevantes, como dietas, hábitos alimentares, alimentos específicos e nutrientes. Cada pergunta foi respondida por ambos os modelos, com e sem raciocínio, e posteriormente avaliada por modelos de linguagem externos.

As tabelas seguintes apresentam os resultados médios atribuídos a cada resposta, bem como a média final por tópico, permitindo uma comparação clara entre as duas abordagens.

- **Dietas:**

Pergunta	Non Reasoning Model	Reasoning Model
What are the benefits of the Mediterranean diet?	7.5	8.8

Pergunta	Non Reasoning Model	Reasoning Model
Is the ketogenic diet safe for diabetics?	6.5	7.8
What is the difference between the paleo diet and the Whole30 diet?	8.5	8.5
Média de Resultados	7.5	8.4

- **Hábitos Alimentares:**

Pergunta	Non Reasoning Model	Reasoning Model
What does mindful eating mean?	6.3	8.5
How to control emotional hunger?	8.0	8.5
What are the effects of eating too fast?	5.8	8.3
Média de Resultados	6.7	8.4

- **Alimentos Específicos:**

Pergunta	Non Reasoning Model	Reasoning Model
What type of fat is avocado rich in?	9.5	9.5
Which vegetables are sources of protein?	7.0	6.8
Which is better for breakfast: oats or bread?	8.0	8.8
Média de Resultados	8.2	8.4

- **Nutrientes:**

Pergunta	Non Reasoning Model	Reasoning Model
What is vitamin D for?	7.8	9.0
What are the symptoms of iron deficiency?	4.5	4.5
What are antioxidants and why are they important?	7.8	8.8
Média de Resultados	6.7	7.4

8.4.2) Modelo Selecionado

Este processo de avaliação permitiu obter uma média de desempenho para cada modelo em diferentes áreas temáticas, de forma a garantir uma comparação justa e imparcial.

Tópico	Non Reasoning Model	Reasoning Model
Dietas	7.5	8.4
Hábitos Alimentares	6.7	8.4
Alimentos Específicos	8.2	8.4
Nutrientes	6.7	7.4
Média Global	7.3	8.2

Apesar de o modelo com raciocínio, *Reasoning Model*, apresentar um tempo de execução superior, os resultados indicaram uma vantagem consistente na qualidade das respostas, o que justificou a sua seleção como modelo preferencial para este agente.

8.5) Configuração do Agente

Nesta secção, apresentam-se os principais parâmetros utilizados na configuração do agente responsável pela geração de respostas. Estes parâmetros foram ajustados de forma a otimizar a qualidade das respostas e o controlo da profundidade e relevância da informação devolvida ao utilizador.

A configuração adotada foi a seguinte:

- **reasoningModel:** true

Esta opção ativa o uso do modelo com capacidade de raciocínio.

- **topK:** 10

Define o número máximo de chunks recuperados com maior correlação com o prompt.

- **targetThreshold:** 0.5

Representa o valor a partir do qual um chunk é considerado fortemente relevante em relação ao prompt.

- **minimumThreshold:** 0.3

Define o valor mínimo de correlação entre um chunk e o prompt para que o conteúdo possa ser considerado como fonte complementar de informação.

- **maxHierarchyLevel:** 3

Indica o nível máximo de hierarquia permitido nos dados a utilizar.

Com excepção dos parâmetros **reasoningModel** e **maxHierarchyLevel**, cuja definição seguiu critérios específicos, o primeiro baseado no processo de avaliação descrito anteriormente, e o segundo determinado pela análise da estrutura hierárquica dos dados disponíveis, os restantes valores foram ajustados de forma empírica, através de sucessivos testes e afinações, com o objectivo de obter melhores resultados observáveis na geração de respostas.

9) Especialista em Suplementação e Fármacos

A automedicação, entendida como o uso de medicamentos ou suplementos sem prescrição nem acompanhamento profissional, é comum entre a população portuguesa, especialmente em contextos de prevenção, via medicamentos de venda livre ou suplementos. No entanto, esta tendência representa riscos substanciais para a saúde pública. Em Portugal, tanto o Infarmed como a Direção-Geral de Alimentação e Veterinária alertam que os suplementos alimentares, apesar da sua ampla disponibilidade e apresentação semelhante a medicamentos, não são sujeitos a avaliações obrigatórias de segurança ou eficácia antes da comercialização [1], [2]. Durante a pandemia de COVID-19, verificou-se um aumento significativo da automedicação com medicamentos não sujeitos a receita médica e suplementos alimentares, sobretudo analgésicos e vitaminas [3]. Esta prática, como refere a Deco Proteste, pode resultar em interações medicamentosas, intoxicações, atrasos no diagnóstico e agravamento da condição clínica [4].

Com a crescente informatização da saúde, ferramentas específicas que ofereçam aconselhamento robusto sobre o uso preventivo de suplementos e medicamentos de venda livre continuam escassas. Modelos de linguagem natural (LLMs) como o GPT-4 têm demonstrado capacidade de resposta comparável a profissionais farmacêuticos humanos. Num estudo recente, o GPT-4 apresentou uma taxa de 75% de respostas completamente precisas a perguntas sobre dosagem e administração de medicamentos, e 55,5% no caso de interações medicamentosas, aproximando-se dos 66,6% obtidos por farmacêuticos humanos [5].

No plano académico, surgem soluções baseadas em LLMs aplicadas à farmacologia. Um exemplo é o modelo desenvolvido por Kim et al. (2024) [6], que combina reconhecimento ótico de caracteres com uma LLM para extrair e analisar interações medicamentosas a partir de documentos clínicos. No contexto comercial, aplicações como a Ada Health [7] oferecem triagem e orientação sintomática baseada em IA, mas raramente integram validação de interações entre suplementos e medicamentos não sujeitos a receita.

9.1) Análise do tema e das fontes de dados

Para a exploração deste tema, foram consideradas diversas fontes de elevada relevância na área da saúde. A seleção teve por base a credibilidade institucional, atualidade e clareza dos conteúdos apresentados, com o objetivo de assegurar uma compreensão abrangente e rigorosa do tema.

Foram exploradas fontes institucionais como o National Institutes of Health (NIH), MedlinePlus (gerido pela Biblioteca Nacional de Medicina dos EUA), Food and Drug Administration (FDA) e European Food Safety Authority (EFSA). Estas entidades disponibilizam conteúdos introdutórios, como explicações gerais sobre suplementos alimentares (benefícios, riscos e formas de utilização), bem como documentos mais específicos sobre substâncias concretas. O NIH, por exemplo, fornece *fact sheets* detalhadas de vários suplementos com dados sobre indicações, efeitos adversos e outras informações pertinentes. Já o MedlinePlus apresenta diretrizes essenciais sobre fármacos. Complementarmente, foi consultada literatura científica atualizada para enriquecer a base de conhecimento, recorrendo a bases de dados académicas como o Europe PMC, PubMed, Semantic Scholar e Google Scholar.

9.2) Metodologia de Recolha e Processamento de Dados

Para construir uma base de conhecimento sólida sobre o tema, foi adotada uma abordagem hierárquica em dois níveis, de forma a garantir simultaneamente a fiabilidade e a profundidade da informação recolhida.

9.2.1) Extração dos dados

No primeiro nível, deu-se prioridade a fontes reconhecidas pela sua autoridade na área da saúde. Foram recolhidos conteúdos sobre suplementos e fármacos a partir de plataformas como o NIH, FDA, MedlinePlus e EFSA. A informação foi inicialmente recolhida manualmente, com foco em conteúdos de carácter geral, e posteriormente complementada com dados específicos sobre determinadas substâncias. Verificou-se que as *fact sheets* do NIH e as *guidelines* do MedlinePlus apresentam uma estrutura padronizada. Com base nesta estrutura, foi elaborada uma lista de suplementos e fármacos relevantes, procedendo-se à extração automatizada dos conteúdos através de técnicas de *web scraping*. Para esta tarefa, foi utilizado um *script* em Python, com o apoio das bibliotecas *BeautifulSoup* e *requests*, permitindo a extração dos textos diretamente das páginas *web* selecionadas. Os dados recolhidos foram posteriormente processados e armazenados. No total, foram reunidos 208 documentos, entre informação geral e fichas técnicas específicas.

O segundo nível teve como objetivo aprofundar a base de conhecimento com conteúdos científicos atualizados. Entre as fontes académicas exploradas, foi selecionado o Europe PMC, pelo seu enquadramento com tema em estudo e pela possibilidade de realizar pesquisas em larga escala, sem restrições significativas de tempo de resposta ou acesso. Foram definidas *queries* específicas centradas na prevenção de doenças e no uso de suplementos e fármacos. Estas pesquisas permitiram a recolha automatizada de milhares de resumos de artigos científicos (abstracts), os quais foram posteriormente processados e integrados na base de dados. Este cruzamento entre informação regulamentar e perspetiva académica permitiu uma abordagem mais abrangente, técnica e cientificamente fundamentada.

Para garantir a relevância e cobertura temática dos dados recolhidos, as *queries* foram formuladas com base na literatura científica mais atual e nas prioridades clínicas e de saúde pública. Algumas das *queries* mais representativas incluem:

- "Multivitamin supplementation" AND "mortality" AND "prevention"
- "Vitamin D" AND "cardiovascular disease" AND "prevention"
- "Supplements for mental health depression anxiety evidence"
- statins AND ("disease prevention" OR "cardiovascular prevention")
- "drug-induced nutrient deficiencies" OR "drug nutrient interactions"

Este cruzamento entre informação regulamentar e perspetiva académica permitiu uma abordagem mais abrangente, técnica e cientificamente fundamentada. A base de conhecimento resultante constitui um alicerce sólido para o desenvolvimento de soluções especializadas na área da suplementação e farmacologia preventiva, com aplicação potencial em contextos clínicos, educativos e tecnológicos.

9.2.2) Processamento dos dados

Os dados recolhidos, especialmente no caso das fontes de nível 1 obtidas através de *web scraping*, foram filtrados para garantir a sua relevância. Foram selecionadas apenas as secções

consideradas pertinentes, excluindo-se partes redundantes ou desnecessárias, como listas de referências bibliográficas ou elementos de navegação da página. O texto extraído foi sujeito a um processo de limpeza e normalização, que incluiu a remoção de citações bibliográficas automáticas, quebras de linha, espaços em excesso e outros elementos não informativos.

Em seguida, aplicou-se uma *pipeline* linguística e semântica baseada na biblioteca *spaCy*, utilizando o modelo `en_core_sci_scibert`, treinado especificamente em literatura científica biomédica. Este modelo permitiu a detecção de entidades nomeadas relevantes, além da realização de uma classificação semântica fundamentada em *matchers* construídos a partir de listas controladas de termos (como doenças, suplementos, fármacos e conceitos médicos), previamente normalizados a partir de ficheiros JSON.

Cada documento foi segmentado em *chunks* com tamanho controlado, respeitando limites impostos pelos modelos de *embeddings* e garantindo que cada segmento mantivesse coerência sintática e semântica. Esta segmentação facilitou as tarefas posteriores de recuperação e análise semântica. Para a representação vetorial dos conteúdos, recorreu-se a um modelo de *embeddings* baseado em *transformers*, o `BAAI/bge-small-en`, adequado para tarefas de recuperação de informação e classificação semântica. O modelo foi carregado via HuggingFace Transformers e utilizado em GPU quando disponível, gerando vetores organizados em *arrays* multidimensionais, onde cada linha corresponde a um *chunk* vetorizado do texto original.

9.2.3) Armazenamento dos dados

Após a geração dos *embeddings*, os vetores foram organizados e inseridos numa solução de indexação vetorial utilizando a API da Pinecone — uma plataforma distribuída otimizada para pesquisa por similaridade em larga escala.

Cada documento, previamente segmentado em *chunks* de texto, manteve o contexto semântico em unidades processáveis. Os segmentos receberam identificadores sequenciais (por exemplo, `_chunk_0`, `_chunk_1`), vinculados a um `doc_id` único. A atribuição deste identificador seguiu uma hierarquia lógica com base nos metadados disponíveis: preferencialmente DOI, seguido de PMID, Semantic Scholar ID, URL ou, caso ausentes, um *hash* determinístico do conteúdo textual. Para assegurar a unicidade e evitar duplicação, foi implementado um mecanismo robusto de deduplicação multiorigem. Este mecanismo verifica, localmente, se o documento já foi processado, consultando o registo `inserted_docs.jsonl`, e remotamente, interrogando a API da Pinecone para a presença do primeiro *chunk* do `doc_id`. Tal procedimento garante que documentos replicados em diferentes fontes (como EuropePMC e PubMed) não sejam indexados múltiplas vezes.

9.3) Benchmarks

Através do *benchmark*, foi possível avaliar qualitativamente a qualidade das respostas geradas pelas *LLMs*. Além do método descrito no capítulo referente a *Role E*, no qual as respostas geradas por diferentes *LLMs* (como *LLama* e *DeepSeek*) foram enviadas para outras três *LLMs* (*ChatGPT*, *Claude* e *Gemini*) com a tarefa de classificá-las de 1 a 10, o grupo também solicitou que cada uma dessas *LLMs* escolhesse a melhor resposta entre as opções e justificasse sua escolha (objeto *Chosen*). Adicionalmente, foi realizada uma avaliação manual (apenas no segundo *benchmark*), na qual três membros do grupo analisaram ambas as respostas e escolheram aquela que consideraram superior (objeto *Manual*). Assim cada ficheiro *json* que possui a informação relativa a cada pergunta, encontra-se com a seguinte [estrutura](#).

9.3.1) Avaliação da Qualidade de Resposta dos Modelos

Os *benchmarks* foram realizados em duas ocasiões distintas, utilizando o mesmo conjunto de perguntas, selecionadas por diferentes critérios que abrangem variados níveis de generalidade e especificidade, desde questões amplas como “Que suplementos devo tomar?”, até conjuntos relacionados, como “O que devo tomar para [condição]?” e “Se tomar em excesso, faz mal?”. Além disso, foram elaboradas perguntas mais específicas que incluíam informações adicionais, como o fato de o usuário ter uma determinada doença (por exemplo: “O que devo tomar para [condição], considerando que tenho a doença Y?”), o que permitiu avaliar a capacidade dos modelos de linguagem a incorporar, neste caso, o contexto clínico na geração das respostas.

9.3.1.1) Primeiro *Benchmark*

Pergunta	<i>Llama</i>	<i>Deepseek</i>	Razão
What are dietary supplements and what are they for?	9.2	9.3	Um pouco mais completa
Are supplements regulated like medications?	8.5	9.2	Mais clara e concisa, mais direta
How do I know if I need a supplement and how long it takes for it to start working?	7.7	9	Mais direta, clara e prática

Abaixo encontra-se a tabela referente à média global das 23 perguntas colocadas às *LLMs*. Podemos verificar que, em geral, o modelo *Deepseek* obteve uma média ligeiramente superior. Outra métrica relevante a considerar é o tempo médio de resposta de cada modelo.

	<i>Llama</i>	<i>Deepseek</i>
Média Global	8.40	8.84

<i>Llama</i>	<i>Deepseek</i>
~4s	~9s

Deste *benchmark* pode concluir-se que o modelo com capacidades de raciocínio (*DeepSeek*) foi consistentemente favorecido pelas restantes *LLMs*, que, na maioria dos casos, consideraram as suas respostas mais completas. Focando no primeiro exemplo da primeira tabela, “*What are dietary supplements and what are they for?*”, apresentamos abaixo as duas respostas geradas pelas diferentes *LLMs*:

- Llama : “Dietary supplements are products intended to add to or “supplement” the diet, and they can come in various forms such as **tablets, capsules, softgels, and liquids**. These products contain dietary ingredients like vitamins, minerals, herbs, and amino acids, which are meant to supplement the diet. The purpose of dietary supplements is to **help improve or maintain overall health, and some can assist in meeting daily requirements of essential nutrients**. For instance, calcium and vitamin D supplements can contribute to building strong bones. However, it is essential to note that dietary supplements **should not replace a healthy and varied diet**. It is crucial to understand that while dietary supplements can be beneficial, they can also involve health risks. Therefore, it is necessary to be informed about the facts before deciding to take a supplement. To obtain more accurate information and guidance on dietary supplements, it is always recommended to consult with healthcare professionals...
- Deepseek - “Dietary supplements are products designed to **complement or enhance the diet by providing additional nutrients or substances that may not be sufficiently**

obtained through food alone. They are available in various forms, **including tablets, capsules, softgels, liquids, gummies, powders, and bars**. These supplements contain dietary ingredients such as vitamins, minerals, herbs, botanicals, amino acids, enzymes, and probiotics, which are intended to support overall health and well-being. The purpose of dietary supplements is to help individuals meet their daily nutritional needs, address potential deficiencies, and promote specific health benefits. For example, calcium and vitamin D supplements are often used to support bone health. However, it is important to note **that supplements should not replace a balanced and varied diet, as they are meant to supplement, not substitute, whole foods**. While dietary supplements can offer health benefits, they may also carry risks. It is crucial to be aware of these potential risks and to consult with healthcare professionals before starting any supplement regimen to ensure safe and appropriate use...

Podemos observar que o *Deepseek* escreve de forma mais detalhada e elaborada. Embora apresente a mesma informação essencial, consegue extrair e acrescentar dados relevantes, proporcionando uma compreensão mais ampla do tema. Por exemplo, ao abordar os suplementos, o *Deepseek* não se limita aos formatos tradicionais, como tabletes, mas também menciona outras formas, como barras, gomas, entre outras. Na nossa opinião, a escrita do *Deepseek* neste primeiro *benchmark* é superior na maioria dos casos, pois gera respostas mais completas e desenvolvidas, respondendo as perguntas com uma maior profundidade. Em ambas as respostas é transmitida a mesma essência, contudo, o *Deepseek* destaca-se por apresentar um conteúdo mais enriquecido, algo que pode ser facilmente verificado nos trechos destacados a *negrito*. Para as restantes perguntas, o mesmo pode ser verificado na maioria dos casos no repositório do *GitHub* do grupo 2([MD - Grupo 2 : Benchmark 1](#)) onde estão disponíveis as restantes avaliações deste primeiro *benchmark*.

No primeiro *benchmark*, verificou-se posteriormente que a presença de documentos repetidos na base de dados do *RAG (Retrieval-Augmented Generation)* (*pinecone*). Este problema comprometeu a diversidade da informação fornecida ao modelo, resultando em respostas menos completas e de qualidade inferior. Esta redundância limitou o valor acrescentado na fase de geração do texto, prejudicando os resultados finais. Para além disso, neste primeiro *benchmark* existiam menos dados disponíveis, o que também afetou a qualidade das respostas a algumas perguntas. Assim, este primeiro *benchmark* foi essencial para identificar possíveis lacunas de dados em determinadas perguntas, especialmente por se tratar de um tema bastante amplo. Além disso, foi útil para detectar erros, conforme descrito no parágrafo anterior. Essa análise contribuiu para enriquecer e melhorar a qualidade das informações utilizadas no segundo *benchmark*.

9.3.1.2) Segundo *Benchmark*

Pergunta	<i>Llama</i>	<i>Deepseek</i>	Razão
What are dietary supplements and what are they for?	8.60	8.30	Um pouco mais completa para um chatbot (<i>DEEPSEEK</i>)
Are supplements regulated like medications?	8.83	6.66	Mais clara e concisa, mais direta, linguagem objetiva e fluída (<i>DEEPSEEK</i>)
How do I know if I need a supplement and how long it takes for it to start working?	6.83	6.16	Detalhada e abrangente (<i>LLama</i>)

Como podemos verificar, embora a média destas respostas tenha diminuído, as respostas do *LLama* passaram a apresentar uma média mais elevada em comparação com as respostas geradas pelo o *Deepseek*. Um comportamento curioso foi observado nas duas primeiras perguntas: as *LLMs* inicialmente classificaram melhor as respostas provenientes do *LLama*, mas, nas avaliação comparativa entre as duas resposta, passaram a preferir as do *DeepSeek*, argumentando que eram mais claras.

	<i>LLama</i>	<i>Deepseek</i>
Média Global	8.56	7.93

Repetindo o processo realizado no primeiro *benchmark* vamos voltar a analisar as respostas geradas pelas as *Llms* para a primeira pergunta:

- *Llama* : “Dietary supplements are products designed to complement the diet by providing essential nutrients. They come in various forms such as **tablets, capsules, and gummies**, and include ingredients like vitamins, minerals, herbs, amino acids, and probiotics. **These supplements aim to support overall health and help meet daily nutritional needs**, such as calcium and vitamin D for bone health. However, **they should not replace a balanced diet**. While they can offer benefits, they also carry potential risks, so safe usage and professional consultation are recommended for personalized advice.
- *Deepseek* : “Dietary supplements are products intended to add to or supplement the diet, and they come in various forms such as **tablets, capsules, soft gels, gel caps, powders, bars, gummies, and liquids**. These supplements are ingested and can contain a range of ingredients including vitamins, minerals, herbs, botanicals, amino acids, and live microbials, also known as probiotics. The primary purpose of dietary supplements **is to help improve or maintain overall health, and they can also assist in meeting daily requirements of essential nutrients**. For instance, calcium and vitamin D can help build strong bones, while fiber can help maintain bowel regularity. It is essential to note that **dietary supplements should not replace a healthy and varied diet, but rather complement it**. Additionally, it is crucial to be aware of the potential risks associated with taking dietary supplements and to use them safely. It is always necessary to contact professionals to obtain more accurate information...

Ao comparar estas respostas com as do primeiro *benchmark*, verifica-se que o conteúdo essencial permanece o mesmo, incluindo até as informações mais específicas que anteriormente eram destacadas pelo *DeepSeek*. Isso torna questionável a avaliação mais baixa atribuída aos outros modelos e levanta dúvidas sobre a razão pela qual o *LLama* apresenta, atualmente, a melhor nota neste *benchmark*. Além disso, verificou-se que, ao enviar a mesma pergunta, especialmente ao *ChatGPT* e ao *Gemini*, as avaliações variavam consideravelmente, o que compromete a consistência e a confiabilidade deste método de *benchmark*. Isso evidencia a necessidade de se considerar outro tipos de abordagens. Por isso, é fundamental reconhecer que a avaliação feita exclusivamente por *LLMs* apresenta limitações e pode não refletir de forma precisa a qualidade das respostas. É necessário complementar esse tipo de *benchmark* com métodos mais objetivos e padronizados que acabam por ser uma avaliação mais manual.

Na avaliação manual, foi escolhida a melhor resposta entre as duas disponíveis. Cada elemento selecionava a resposta que preferia, podendo justificar sua escolha. Considerando os dois textos gerados para a primeira pergunta, os resultados da avaliação manual foram os seguintes:

- Dois dos três elementos optaram pela resposta do *Deepseek* apenas por esta possui mais informação enquanto que o outro elemento preferiu a resposta proveniente do *Llama* devido a considerar a resposta mais coesa gramaticalmente.

As restantes das avaliações deste *benchmark* podem ser visualizadas nesta hiperligação [MD - Grupo 2 : Benchmark 2](#). É importante referir que o tempo médio que as LLMs demoram a responder se manteve muito semelhante ao do *benchmark* anterior, motivo pelo qual o grupo optou por não incluir novamente a tabela correspondente.

9.3.2) Modelo Selecionado

Dos dois *benchmarks* foi possível observar dois pontos bastante importantes para:

- Em grande parte, a maior qualidade dos dados fornecidos pelo *RAG* influenciou significativamente os resultados das avaliações das LLMs. Quando os *chunks* continham mais informações relevantes, o *LLama* foi capaz de sintetizar melhor as respostas, conforme solicitado no conjunto de regras do *prompt*, enquanto os textos gerados pelo o *deepseek* tendiam a aumentar considerativamente.
- Em contrapartida, o *DeepSeek* tende, por vezes, a ignorar algumas das instruções definidas no *prompt* inicial. Por exemplo, apesar de ser explicitamente solicitado que a resposta seja redigida unicamente em parágrafos, sem recorrer a *markdown* ou enumerações, conforme a regra “*Write only in paragraphs and never use languages like markdown. Never enumerate too, only write in paragraphs*”, o *DeepSeek* desrespeita ocasionalmente estas orientações, utilizando *markdown* ou pontuação para enumerar itens. Outro problema observado é que o *DeepSeek* tende a considerar grande parte da informação fornecida como relevante, o que resulta num aumento do comprimento do texto. Isto contraria, em certa medida, a instrução dada no *prompt*, “*Write the text in such a way that it’s not too long but contains most of the information you want to convey*”, que procura um equilíbrio entre concisão e riqueza informativa.

Para além destes dois aspetos que tornam o *DeepSeek* um modelo mais imprevisível, este apresenta também um tempo médio de resposta mais do dobro em relação ao *LLama*. Perante estes fatores, o grupo optou por selecionar o **LLama** como a melhor opção.

9.4) Configuração do Agente

- **reasoningModel**: Como já foi referido anteriormente o modelo escolhido foi o *Llama* (**false**)
- **topK**: Para escolher quantos *chunks* seriam utilizados para serem passado para a LLM o grupo realizou testes com 3, 4 e 5 *chunks*. A diferença entre 4 e 5 *chunks* normalmente não é muito perceptível, mas pode tornar-se mais evidente nos textos quando os scores do *RAG* são mais baixos (abaixo dos 30%). Nesse sentido o grupo optou por escolher **5** *chunks* para nesses casos em específico ser possível, obter a melhor resposta possível. De notar com mais *chunks* as respostas tornariam-se mais longas, algo que nem sempre é bom.
- **targetThreshold** e **minimumThreshold**: O grupo reparou que mensagens geradas de *chunks* que tinham um score de **0.5** produziam respostas bastante completas o que levou a escolha deste valor para o **targetThreshold**. O valor **0.25** foi escolhido para o **minimumThreshold** com base em testes realizados, nos quais era pedido que o modelo resumisse o texto enviado anteriormente. Durante esses testes, observou-se que as vezes os *prompts* conseguiam localizar *chunks* com scores inferiores a 0.25 (inúteis para a tarefa pedida).
- **maxHierarchyLevel**: Como só existem 2ª hierarquias definidas o valor deste campo encontra-se a **2**.

10) Especialista em Atividade Física

10.1) Análise do Tema e das Fontes de Dados Relevantes

O estudo da atividade física e dos seus impactos na saúde integra uma abordagem multidisciplinar, englobando áreas como fisiologia, psicologia, saúde pública e ciências do desporto. Os efeitos benéficos da prática regular de exercício físico estão bem documentados, abrangendo:

- Melhoria da função cardiovascular, com redução do risco de hipertensão, enfarte do miocárdio e acidente vascular cerebral;
- Aumento da densidade mineral óssea, prevenindo condições como a osteoporose;
- Desenvolvimento da força e resistência muscular, com impacto direto na funcionalidade e autonomia ao longo do ciclo de vida;
- Melhorias significativas na saúde mental, incluindo a redução dos sintomas de ansiedade e depressão, o aumento da autoestima e o reforço das funções cognitivas;
- Regulação do peso corporal e da composição corporal, contribuindo para a prevenção e gestão de doenças metabólicas como a diabetes tipo 2;
- Reforço do sistema imunitário e melhoria dos padrões de sono.

A relevância de compreender estas relações é particularmente importante na formulação de políticas de saúde pública e em estratégias de intervenção personalizadas, adaptadas a diferentes contextos sociodemográficos.

As fontes de dados utilizadas para esta análise incluíram artigos científicos obtidos a partir de bases de dados académicas e institucionais de alto impacto, como:

- **PubMed e National Library of Medicine** — para revisões sistemáticas e estudos clínicos controlados;
- **Google Scholar** — para uma pesquisa exploratória ampla de literatura científica multidisciplinar;
- **Nature, Springer, BMJ Journals, PLOS, Frontiers e Cambridge** — para artigos de investigação recentes, submetidos a revisão por pares;
- **Organizações internacionais** como a **Organização Mundial da Saúde (OMS)** e os **Centers for Disease Control and Prevention (CDC)** — fontes fundamentais para a obtenção de diretrizes baseadas em evidência sobre a frequência, intensidade, duração e tipo de exercício físico recomendado para diferentes grupos populacionais.

Estas fontes foram criteriosamente selecionadas tendo em conta o seu rigor científico, atualidade da informação e relevância para o contexto europeu.

10.2) Metodologia de Recolha e Processamento de Dados

10.2.1) Estratégia de Web Scraping

A recolha de dados foi realizada através de uma abordagem automatizada de **web scraping**, com o objetivo de extrair conteúdos relevantes de forma sistemática e eficiente. Esta estratégia permitiu ultrapassar limitações associadas à pesquisa manual, como a dispersão da informação e a heterogeneidade dos formatos. Foram definidos critérios de pesquisa com base em palavras-chave específicas (e.g., **physical activity, exercise benefits, public health guidelines**) e parâmetros temporais (publicações entre 2018 e 2024), de modo a garantir a atualidade dos dados.

Além disso, foram implementados filtros semânticos e de validação da fonte, assegurando que apenas publicações indexadas e de relevância reconhecida fossem incluídas. Esta metodologia permitiu construir um corpo de evidência robusto e fiável.

10.2.2) Ferramentas de Extração

Na execução técnica do **web scraping**, foram utilizadas bibliotecas de Python altamente eficazes:

- **BeautifulSoup** — para a análise e navegação do conteúdo HTML, facilitando a extração de metadados (autores, data de publicação, título, palavras-chave);
- **Requests** — para a realização de chamadas HTTP e acesso a páginas específicas de repositórios de artigos científicos;
- **PyPDF2** e **pdfminer.six** — para o processamento de documentos em formato PDF, permitindo extrair o conteúdo textual integral de artigos que não estavam disponíveis em formato HTML;
- **Pandas** — para organização, limpeza e estruturação dos dados extraídos, facilitando posteriores análises quantitativas e qualitativas.

O processo foi desenvolvido respeitando rigorosamente as políticas de uso de cada plataforma, incluindo verificações de **robots.txt**, uso de **user agents** identificáveis e limitação de chamadas por minuto para evitar sobrecarga dos servidores.

A combinação destas ferramentas e práticas assegurou uma recolha de dados sistemática, ética e tecnicamente sustentada, apta para suportar uma análise aprofundada sobre o tema da atividade física e saúde.

10.2.3) Definição de Keywords por Domínio

As **keywords** foram definidas estrategicamente para abranger os domínios fundamentais da investigação em atividade física, assegurando uma cobertura abrangente das áreas com maior relevância científica e clínica. A definição destas palavras-chave permitiu afinar a recolha de dados e garantir a pertinência dos conteúdos extraídos.

- **physical_activity**: abrange os benefícios gerais da prática regular de atividade física, com ênfase na sua influência positiva sobre a saúde cardiovascular, metabólica e mental. Esta keyword foi central para a identificação de estudos longitudinais e meta-análises em saúde pública.
- **aerobic_exercise**: foca-se nos efeitos fisiológicos do exercício aeróbico, incluindo a melhoria da capacidade cardiorrespiratória, o aumento da resistência e o papel na regulação do peso corporal e da sensibilidade à insulina.
- **strength_training**: centra-se nos impactos do treino de força e resistência muscular, nomeadamente no aumento da massa magra, na densidade mineral óssea, na melhoria da funcionalidade motora e na prevenção de quedas e lesões, especialmente em populações envelhecidas.
- **flexibility_exercises**: abrange práticas como alongamentos estáticos, dinâmicos e técnicas como yoga e pilates, analisando a sua importância na manutenção da mobilidade articular, na prevenção de rigidez muscular e na promoção de uma postura funcional.

- **exercise_by_age_group**: direcionada à identificação de diretrizes e recomendações específicas por faixa etária (crianças, adolescentes, adultos e idosos), tendo em conta as capacidades fisiológicas, os riscos específicos e os benefícios esperados em cada grupo.
- **mental_health_and_exercise**: explora a ligação entre a prática de exercício físico e a saúde mental, com foco na evidência científica sobre a redução de sintomas de ansiedade, depressão, stress crónico e declínio cognitivo. Inclui também estudos sobre os mecanismos neurobiológicos envolvidos, como a libertação de endorfinas e a neurogénese.

Esta segmentação semântica das keywords facilitou uma extração orientada e temática de conteúdos, possibilitando uma análise posterior mais estruturada e segmentada por área de interesse.

10.2.4) Chunking de Dados

A fase final da preparação dos dados consistiu na segmentação textual em unidades menores (**chunks**), um passo essencial para posterior análise semântica, vetorização e integração em modelos de aprendizagem automática. Esta tarefa visou preservar a coesão temática e o contexto técnico dos documentos originais, mesmo após a sua fragmentação.

Foram utilizadas ferramentas de processamento de linguagem natural como **SpaCy** e o **tokenizer** da biblioteca **HuggingFace Transformers**, permitindo a divisão dos textos com base em estruturas linguísticas (frases, parágrafos e secções semânticas), em vez de divisões arbitrárias por número de caracteres. Isto assegurou que cada chunk mantivesse a integridade dos conceitos científicos abordados.

Este processo de **chunking** teve múltiplas finalidades:

- Facilitar a criação de embeddings vetoriais semanticamente ricos;
- Permitir a indexação temática eficiente para mecanismos de recuperação de informação;
- Preparar os dados para tarefas subsequentes de classificação, sumarização e question answering com modelos de linguagem avançados.

A segmentação foi ajustada de forma adaptativa, tendo em conta a densidade informacional e a complexidade técnica de cada fonte textual, garantindo uma cobertura representativa e tecnicamente robusta dos principais tópicos associados à atividade física e à promoção da saúde.

10.2.5) Hierarquia

Hierarquia	Tipo de Fonte	Exemplos	Justificação
1	Bases de dados e bibliotecas oficiais de saúde e medicina	PubMed, National Library of Medicine (NLM)	Fontes oficiais mantidas por organismos públicos, com revisão rigorosa e elevada confiança científica.
2	Publicações de editoras científicas de referência	Nature, Springer, BMJ Journals, Frontiers, Cambridge, Plos	Revistas peer-reviewed de elevado prestígio, com elevado fator de impacto e rigor científico.
3	Motores de busca científica, agregadores	Google Scholar	Inclui publicações de qualidade variável, requerendo verificação adicional da fiabilidade.

10.3) *Benchmarks* de Modelos

Esta secção justifica a escolha do modelo com capacidade de raciocínio (Reasoning Model) em detrimento do modelo sem raciocínio (Non Reasoning Model). A decisão baseou-se numa avaliação comparativa da qualidade das respostas de ambos os modelos. Para isso, utilizaram-se modelos de linguagem externos — ChatGPT-4.0, Claude Sonnet 4, Gemini 2.5 Flash e Grok 3 — que atribuíram pontuações de 1 a 10 às respostas, com base em critérios como clareza, precisão, completude e relevância.

10.3.1) Avaliação da Qualidade de Resposta dos Modelos

Para avaliar o desempenho dos modelos, foram elaboradas perguntas distribuídas por diferentes tópicos relevantes e presentes nas nossas fontes de dados. Cada pergunta foi respondida por ambos os modelos — com e sem raciocínio — e posteriormente avaliada por modelos de linguagem externos.

Pergunta	Non Reasoning Model	Reasoning Model
What are the key physical activity recommendations guidelines for adults?	8	9.3
How much daily physical activity is recommended for children and adolescents?	7.5	8
What are some low-impact cardio options for someone with joint issues?	9	9.3
What types of exercise do you recommend for post-stroke rehabilitation?	7	8.5

Pergunta	Non Reasoning Model	Reasoning Model
What are the benefits of aerobic versus resistance training for a individual with depression?	8.3	8
What is reccomended exercise plan for children with asthma?	7.5	8.3
How does resistance training benefit cognitive function?	8	9
What is the difference between physical activity and physical exercise?	9.1	9.5
Why is reducing sedentary time important, even if I'm already active?	8.8	8.5
Consider that i am overweight, does exercise really help with anxiety? If so, what kind of exercise is best for me?	7.7	8
Média de Resultados	8.1	8.6

10.3.2) Modelo Selecionado

Apesar do maior tempo de execução associado ao **Reasoning Model** (deepseek-ai/DeepSeek-R1-Distill-Llama-70B-free), os resultados obtidos demonstraram uma melhoria consistente na qualidade das respostas. Esta vantagem justifica a sua adoção como modelo preferencial na configuração do agente.

10.4) Configuração do Agente

Nesta secção, descreve-se a configuração final do agente especializado em atividade física, com destaque para os parâmetros fundamentais que orientam o seu comportamento na fase de geração de respostas. A definição destes parâmetros teve como principal objetivo maximizar a qualidade, relevância e clareza da informação apresentada ao utilizador, assegurando simultaneamente a coerência contextual e a precisão científica.

A configuração adotada baseou-se em princípios de recuperação semântica, raciocínio estruturado e hierarquização da informação, e foi otimizada com base em testes iterativos. Os principais parâmetros foram os seguintes:

- **reasoningModel: true** Ativa a utilização de um modelo com capacidades avançadas de raciocínio, permitindo ao agente construir respostas logicamente encadeadas e semanticamente ricas, com maior capacidade de inferência sobre relações implícitas nos dados.
- **topK: 5** Define o número máximo de segmentos informativos (**chunks**) a recuperar por consulta, com base no grau de similaridade semântica com o prompt. Este valor foi selecionado para equilibrar profundidade e concisão na composição da resposta, evitando redundância informacional.
- **minimumThreshold: 0.3** Representa o valor mínimo de similaridade (tipicamente baseado em cosine similarity de embeddings) entre um chunk e o prompt, necessário para que

esse chunk seja considerado elegível para inclusão no contexto. Este limiar filtra conteúdos irrelevantes, promovendo a pertinência dos dados utilizados.

- **targetThreshold: 0.59** Define o valor de corte a partir do qual um chunk é considerado altamente relevante e prioritário. Chunks que atingem ou superam este limiar são incorporados automaticamente na construção da resposta, garantindo uma base factual sólida.
- **maxHierarchyLevel: 3** Estabelece o nível máximo de profundidade hierárquica a explorar dentro das estruturas informacionais (ex.: capítulos, subsecções, listas aninhadas) dos documentos fonte. Este parâmetro limita a granularidade da análise, assegurando a coesão do conteúdo sem dispersar a resposta em detalhes excessivos.

Importa referir que, com exceção do parâmetro **maxHierarchyLevel** — definido com base na análise estrutural das fontes documentais utilizadas — os restantes parâmetros foram ajustados de forma empírica. Este processo envolveu múltiplas iterações experimentais, com testes de recuperação e validação manual da qualidade das respostas, de modo a identificar a combinação de valores que melhor equilibra os seguintes pontos:

- Relevância semântica dos dados recuperados;
- Profundidade analítica das respostas;
- Clareza e acessibilidade da informação para o utilizador final.

Esta configuração, embora otimizada para o domínio da atividade física e saúde, mantém flexibilidade para ser ajustada a outros contextos temáticos, caso se pretenda reorientar o agente para novas áreas de especialização.

11) Especialista em Hábitos Diários

11.1) Análise do tema e das fontes de dados relevantes

O estudo de hábitos relacionados ao sono, jornadas de trabalho, consumo de substâncias como tabaco, álcool e drogas, e práticas ergonômicas é igualmente essencial para a saúde pública. Estes fatores comportamentais estão fortemente associados a diversas doenças crônicas, distúrbios mentais, acidentes laborais e redução da qualidade de vida. Compreender esses padrões permite identificar não apenas a frequência ou intensidade de tais comportamentos, mas também os determinantes sociais, culturais e ambientais que os influenciam.

Analisar estes hábitos é fundamental para desenvolver intervenções eficazes que promovam estilos de vida mais saudáveis, previnam doenças e melhorem o bem-estar geral da população. Por exemplo, a privação de sono está associada a maior risco de hipertensão, depressão e obesidade; já o excesso de trabalho pode provocar esgotamento físico e mental. O consumo de substâncias continua a ser uma das principais causas evitáveis de morbidade e mortalidade, enquanto a má ergonomia está ligada a problemas músculo-esqueléticos e lesões ocupacionais.

Para além das fontes mencionadas anteriormente pela Role C, foram incluídas outras mais direcionadas a estes temas. Um exemplo é o site da National Sleep Foundation (sleepfoundation.org), que disponibiliza informação baseada em evidências sobre a importância do sono adequado para a saúde física e mental. Este recurso é útil tanto para a população em geral como para profissionais de saúde. Já em termos de ergonomia e ambiente de trabalho, destaca-se a Occupational Safety and Health Administration (OSHA.gov), que publica diretrizes e estudos sobre prevenção de lesões relacionadas ao trabalho, abordando práticas recomendadas e normas de segurança.

11.2) Extração de Conhecimento Externo

11.2.1) Extração Manual

Na fase de Extração Manual, o grupo valorizou a confiança e a credibilidade das informações recolhidas. Este tipo de extração ocupa o topo da hierarquia de fiabilidade dos dados, dado que envolve uma seleção humana criteriosa de fontes e conteúdos. Foram priorizadas fontes oficiais e altamente respeitadas, tais como sites governamentais (.gov), a Organização Mundial da Saúde, a National Sleep Foundation e a Agência Europeia para a Segurança e Saúde no Trabalho. A ênfase esteve na atualidade dos dados, na pertinência do conteúdo e na fidedignidade das fontes utilizadas. Assim, recolhemos a olho e com critério dezenas de excertos relevantes para o nosso tema.

11.2.2) Recolha Automatizada de Dados

A etapa seguinte, de recolha de dados, recorreu a técnicas de extração digital através de palavras-chave específicas relacionadas com: álcool e drogas, higiene, trabalho, ergonomia, sono e tabaco. As principais fontes foram bases de dados científicas, como o **PubMed** e o **Google Scholar**.

O PubMed, classificado como segundo nível na hierarquia de confiança por apresentar contextos mais pequenos e com informação resumida sob a forma de abstracts, disponibilizou milhares de resumos de artigos científicos. Já o Google Scholar, por apresentar papers completos e, conseqüentemente, conter algum ruído e informação irrelevante, foi considerado de terceiro nível. Ainda assim, permitiu o acesso a dezenas de artigos académicos em formato PDF. Esta

abordagem visou complementar a extração manual com uma maior abrangência e volume de informação, sem comprometer totalmente a qualidade das fontes.

11.2.3) Processamento de Dados

Relativamente ao processamento de dados, foram aplicadas várias etapas de limpeza e preparação para posterior análise e embedding. Isto incluiu a remoção de cabeçalhos, números de página, referências bibliográficas, URLs, emails, símbolos invulgares e espaços desnecessários. Esta normalização permitiu uniformizar os textos recolhidos e melhorar a sua qualidade semântica.

Numa primeira abordagem, o grupo procurou resumir artigos extensos com recurso a modelos pré-treinados, numa tentativa de extrair apenas o essencial. No entanto, os resultados não justificaram o aumento de complexidade e custo computacional envolvido, pelo que optámos por uma abordagem mais simples e direta, mas não menos eficaz.

11.2.4) Chunking de Dados

Por fim, o processo de chunking — ou seja, a divisão dos textos em unidades menores — foi feito utilizando ferramentas como o modelo SpaCy e o tokenizer da HuggingFace. Os textos foram divididos em chunks de tamanho controlado, mantendo o contexto necessário para uma boa representação vetorial. No total, foram gerados cerca de **9000** chunks.

11.3) *Benchmarks*

Nesta secção, iremos analisar o desempenho dos modelos avaliados, considerando dois aspetos principais: a qualidade das respostas geradas e o tempo de execução. Para isso, foram realizadas avaliações detalhadas em diferentes áreas temáticas, comparando os modelos com raciocínio (*Reasoning Model*) e sem raciocínio (*Non Reasoning Model*).

Além de apresentar os resultados quantitativos, também discutiremos o impacto do tempo de resposta na escolha do modelo mais adequado. No final, será feita uma conclusão sobre qual modelo se mostrou mais capaz de equilibrar qualidade e eficiência.

11.3.1) Avaliação da Qualidade de Resposta dos Modelos

Foram realizadas avaliações de desempenho com base em três modelos de linguagem: ChatGPT-4.0, Claude Sonnet 4 e Gemini 2.5 Pro, com o objetivo de comparar a qualidade das respostas geradas e ter alguma amplitude de testes.

Para cada modelo, foi atribuída uma nota de 1 a 10, avaliando a qualidade das respostas fornecidas. Estas notas foram utilizadas para calcular a média ponderada, permitindo uma análise comparativa mais detalhada entre os diferentes modelos e configurações.

Além disso, iremos abordar cada subtópico de forma individual, avaliando os resultados específicos de cada área. No final, os resultados serão justificados e consolidados, permitindo uma visão global do desempenho dos modelos em diferentes contextos.

- **Álcool e Drogas:**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
What are the long-term health effects of regular alcohol use?	8.3	9

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
How do recreational drugs affect brain and liver function?	8.3	8.7
What are the risks of mixing alcohol with prescription or illegal drugs?	9.3	9.6
Média de Resultados	8.6	9.1

- **Higiene**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
How does personal hygiene help prevent infectious diseases?	9.3	10
What health problems are linked to poor hygiene?	9.3	9.3
Média de Resultados	9.3	9.7

- **Trabalho**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
What are the health consequences of working over 60 hours per week?	9.3	8.7
How does night shift work affect sleep and hormone regulation?	9	9.3
What are effective strategies to reduce health risks in shift workers?	8.3	8.7
Média de Resultados	8.9	8.9

- **Ergonomia**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
How does poor office ergonomics contribute to chronic back pain?	8.7	9.7
Why is good posture important for long-term musculoskeletal health?	7.7	9.3
How can remote workers improve their home ergonomics?	9.3	8.7
Média de Resultados	8.6	9.2

- **Sono**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
How does poor sleep quality impact mental and physical health?	9.3	10
Why is sleep especially important for teenagers and young adults?	8.3	9.3
How does irregular sleep affect the body's circadian rhythm?	9.3	9
What are the benefits of maintaining a consistent sleep routine?	8	7.7
Média de Resultados	8.7	9

- **Tabaco**

Pergunta	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
How does tobacco smoke damage the lungs over time?	9	7.7
What are the cardiovascular risks of regular nicotine use?	9.7	9.3
Why is tobacco considered a leading preventable cause of death?	9.3	9
What are the effects of passive smoking on children?	9.3	8.3
How does quitting smoking improve health within the first year?	9.7	9
Média de Resultados	9.4	8.7

Os resultados finais mostram que, de forma geral, o modelo com raciocínio supera ligeiramente o modelo sem raciocínio. Em temas como “Higiene” e “Ergonomia”, o *Reasoning Model* apresentou um desempenho significativamente superior, refletindo a sua capacidade de lidar com questões mais complexas, como podemos ver pela tabela seguinte.

Tema	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
Álcool e Drogas	8.6	9,1
Higiene	9.3	9.7
Trabalho	8.9	8.9
Ergonomia	8.6	9.2
Sono	8.7	9
Tabaco	9.4	8.7

Tema	<i>Non Reasoning Model</i>	<i>Reasoning Model</i>
Média Global	8.9	9.1

Embora em alguns casos, como “Trabalho”, ambos os modelos tenham obtido desempenhos semelhantes, a média geral dos resultados confirma a vantagem do *Reasoning Model* (9.1) em relação ao *Non Reasoning Model* (8.9). Esta diferença não é significativa em termos de avaliação puramente quantitativa. No entanto, ela reflete uma ligeira tendência do modelo com raciocínio (*Reasoning Model*) em oferecer respostas mais fundamentadas em determinados contextos.

11.3.2) Tempo de Execução

Em termos de tempo de execução, o modelo com raciocínio (*Reasoning Model*) apresenta uma média de 11 segundos para gerar uma resposta, enquanto o modelo sem raciocínio (*Non Reasoning Model*) é significativamente mais rápido, com uma média de apenas 5 segundos. Esta diferença é bastante considerável, o modelo com raciocínio demora, em média, 2.2 vezes mais tempo a gerar uma resposta.

11.3.3) Modelo Escolhido

O modelo escolhido foi o *Non Reasoning Model*. Apesar de ambos os modelos apresentarem um desempenho muito bom, a pequena melhoria observada no *Reasoning Model* não justifica o tempo de execução significativamente maior, que é mais de o dobro do *Non Reasoning Model*. Assim, o *Non Reasoning Model* foi considerado a melhor opção, equilibrando qualidade e eficiência.

11.4) Configuração do Agente

A configuração do modelo foi projetada para ajustar o comportamento do agente de acordo com os requisitos específicos do sistema. Os parâmetros definidos permitem controlar aspectos importantes do processamento e da recuperação de dados. Abaixo estão os principais parâmetros utilizados:

- **reasoningModel:** Define se o modelo de raciocínio será utilizado. Neste caso, está configurado como `false`, indicando que o agente não utilizará raciocínio avançado.
- **topK:** Define o número máximo de *chunks* a serem considerados durante a recuperação de dados. Neste caso, está configurado para 5.
- **targetThreshold:** Determina o *score* ideal para a seleção de resultados relevantes, configurado como 0.6.
- **minimumThreshold:** Especifica o *score* mínimo aceitável para os resultados, configurado como 0.3.
- **maxHierarchyLevel:** Limita o número de níveis hierárquicos a serem explorados durante a recuperação de dados, configurado como 3.

Esses parâmetros foram ajustados para garantir que o agente opere de forma eficiente, equilibrando qualidade e desempenho no processamento das solicitações.

12) Especialista em Check-ups/acompanhamento médico de medicina preventiva

12.1) Exploração do Tema e Identificação de Fontes Relevantes

A área da medicina preventiva, e em particular a realização periódica de check-ups médicos, assume um papel central na promoção da saúde e na detecção precoce de doenças. Estes exames visam identificar fatores de risco e alterações fisiológicas antes que se manifestem sintomas clínicos, permitindo intervenções atempadas e eficazes. A sua importância está amplamente documentada na literatura médica, sendo considerada uma das estratégias mais custo-efetivas para reduzir a carga de doenças crônicas como diabetes, hipertensão, cancro e doenças cardiovasculares.

Além de melhorar os prognósticos individuais, o acompanhamento médico preventivo contribui significativamente para a sustentabilidade dos sistemas de saúde, ao evitar tratamentos de maior complexidade e custo em fases avançadas da doença. A adesão a programas de rastreio, como rastreios oncológicos (mamografias, colonoscopias, etc.), avaliação dos níveis de colesterol, controlo da tensão arterial e exames de rotina adaptados à faixa etária e sexo, revela-se essencial para a vigilância epidemiológica e melhoria da qualidade de vida da população.

12.2) Extração de Conhecimento Externo

12.2.1) Extração Manual de Informação Confiável

Na fase de Extração Manual, foi dada primazia à qualidade, atualidade e credibilidade das fontes. A equipa focou-se em entidades oficiais, sites institucionais e organizações de saúde pública reconhecidas internacionalmente, como o CDC e a OMS. Esta abordagem garantiu que os conteúdos recolhidos refletiam o consenso científico atual e as melhores práticas clínicas. Foram selecionados documentos técnicos, guidelines, relatórios de saúde e páginas institucionais com informação educativa e epidemiológica sobre exames de rotina e rastreios recomendados.

12.2.2) Recolha Automática de Dados

Na componente automatizada, foram utilizadas palavras-chave como “preventive check-ups”, “screening guidelines”, “health assessment protocols” e “early disease detection” em bases de dados científicas, nomeadamente o PubMed e o Google Scholar. O PubMed revelou-se especialmente útil na recolha de abstracts de estudos recentes sobre a eficácia dos check-ups na prevenção de doenças. Já o Google Scholar permitiu o acesso a documentos mais extensos e revisões sistemáticas, embora com maior variação na qualidade da informação.

Estas ferramentas complementaram a extração manual, oferecendo maior volume e diversidade de perspectivas. Foi feita uma triagem criteriosa dos resultados, priorizando fontes peer-reviewed e artigos com elevada relevância clínica.

12.2.3) Tratamento e Preparação dos Dados Recolhidos

Após a recolha, os dados passaram por um processo de limpeza e preparação. Foram eliminados elementos desnecessários como referências, links, cabeçalhos técnicos e formatações erráticas. Este passo teve como objetivo normalizar os textos e garantir uma consistência semântica adequada para posterior análise automática.

12.2.4) Chunking de Dados

A fase final consistiu na segmentação dos textos em unidades menores (chunks), preservando a coesão temática. Ferramentas como SpaCy e o tokenizer da HuggingFace foram utilizadas para dividir os textos de forma a manter contextos clínicos úteis para tarefas futuras de representação vetorial e embedding. Foram produzidos aproximadamente 850 chunks, garantindo cobertura abrangente dos principais tópicos relacionados com medicina preventiva e check-ups.

12.3) Context Prompt Usado

Act as a medical prevention and public health expert.

Use the following rules to answer questions about medical check-ups and preventive health-care:

If scientific article excerpts are provided, base your answer only on those excerpts. If the user asks to simplify, summarize, or explain a previous answer, and no new content is provided, review the assistant’s previous messages included in the “UserHistory” section below. These represent prior assistant responses in the conversation. If neither article excerpts nor relevant conversation history is available, reply: “There is not enough information to answer this question completely.”

Your answer should consider, when applicable, factors such as: Medical check-ups, follow-up appointments, screening programs, early detection of diseases, monitoring of risk factors, adherence to treatment, and continuity of preventive care.

Write in clear, objective, fluid English using complete sentences and well-structured paragraphs. Integrate all relevant information into cohesive paragraphs. Do not summarize article by article. If explicit information about the user is given, only use that information in the answer if it is relevant to the question. Do not repeat similar ideas using different wording. Each paragraph should add a unique perspective. When referring to articles, cite them using their title and year. Only include details about the user if clearly relevant to the question. Do not evaluate or comment on the articles themselves. Ignore any article that is unrelated to the question — do not mention them. Do not use lists, bullet points, or numbered items in your response. Write plaintext paragraphs (never use markdown or any other formatting language).

12.4) Benchmarks

Nesta secção, analisamos o desempenho dos modelos de linguagem utilizados no nosso sistema, com foco na qualidade das respostas geradas. Para tal, avalíamos 20 perguntas diferentes dentro do tema dos check-ups e da medicina preventiva, testadas em três modelos: ChatGPT-4.0, Claude Sonnet 4 e Gemini 2.5 Pro. Cada pergunta foi avaliada em duas variantes: com e sem o **reasoning model** ativado.

12.4.1) Avaliação de Qualidade: Exemplos Qualitativos

Abaixo apresentamos dois exemplos de perguntas e as respetivas respostas geradas por cada modelo com e sem reasoning model. Foram escolhidas por ilustrarem diferenças subtis na profundidade e organização da resposta.

12.4.1.1) Exemplo 1

Pergunta: What are the benefits of regular medical check-ups for disease prevention?

Resposta sem Reasoning Model: Regular medical check-ups play a crucial role in disease prevention by providing numerous benefits. Foremost, they help establish a person’s medical

history, which is essential in case a problem arises that needs attention. These evaluations often include important screenings that can be used to detect problems early, allowing for timely intervention and treatment. Routine check-ups also enable healthcare providers to offer suggestions for lifestyle improvements, which can significantly contribute to disease prevention. Furthermore, early detection and screening for diseases such as cancer, diabetes, and heart disease can increase the chances of successful treatment and improve outcomes. By detecting and treating conditions early, healthcare providers can also help prevent the development of more serious health problems.

In addition to early detection and screening, routine medical check-ups can help identify and address risk factors that contribute to mortality. According to research, tobacco, poor diet, physical inactivity, and misuse of alcohol are responsible for a significant number of deaths annually in the United States. Many of these deaths could have been prevented through routine check-ups and medical care. Immunization is also a key preventive healthcare strategy, helping to protect individuals from a range of infectious diseases. By undergoing routine medical reviews, individuals can take a proactive approach to their health, stopping problems before they start. Many health insurance companies cover annual visits, making these available at no additional cost to policy holders, which can help reduce long-term medical costs.

Overall, regular medical check-ups are a vital component of preventive care, enabling individuals to take control of their health and well-being. By establishing a medical history, detecting problems early, and addressing risk factors, routine check-ups can help prevent the development of serious health problems and improve overall health outcomes. As emphasized by healthcare providers, preventive check-ups are one key to living a healthy life, and individuals should take advantage of available resources, such as annual visits covered by health insurance companies, to prioritize their health and well-being.

Resposta com Reasoning Model: Regular medical check-ups offer numerous benefits for disease prevention, each contributing uniquely to overall health and wellness.

Firstly, these check-ups facilitate the early detection of diseases such as cancer, diabetes, and heart disease. Early detection significantly improves treatment success rates and outcomes, as highlighted in “The Importance of Preventive Healthcare Strategies for Better Health and Wellness.” This early intervention can prevent the progression of conditions into more severe stages.

Secondly, routine check-ups establish a comprehensive medical history, which serves as a crucial reference point for future health issues. As noted in “The Importance of Preventive Check-ups,” this history is essential for tailored care and timely interventions.

Thirdly, check-ups provide opportunities for healthcare providers to offer personalized lifestyle advice. Discussions about diet, exercise, and substance use can lead to meaningful behavioral changes, preventing diseases before they develop.

Additionally, immunization is a key component of preventive care. Vaccines protect against infectious diseases, as emphasized in the same article, and have been instrumental in reducing the incidence of many such diseases.

Lastly, many insurance plans cover annual check-ups, making preventive care accessible and cost-effective. This financial incentive, mentioned in the article referencing the New England

Journal of Medicine study, encourages regular health maintenance, potentially reducing long-term medical costs.

In summary, regular medical check-ups are a cornerstone of preventive healthcare, offering early detection, historical health insights, lifestyle guidance, immunization, and cost-effectiveness, all of which contribute to better health outcomes.

12.4.1.2) Exemplo 2

Pergunta: In what ways do check-ups contribute to the early identification of risk factors?

Resposta sem Reasoning Model: Check-ups play a crucial role in the early identification of risk factors by providing an opportunity for healthcare providers to evaluate an individual's medical history, lifestyle choices, and genetic predispositions. During a check-up, discussions with the doctor and/or nurse about family history, past surgeries, allergies, tobacco, alcohol or illicit drug use, and any herbal or over-the-counter medicines or vitamins can help identify potential risk factors. Additionally, screenings and blood tests can detect problems early, allowing for prompt intervention and prevention of more serious health issues. The early detection and screening of diseases such as cancer, diabetes, and heart disease can increase the chances of successful treatment and improve outcomes, highlighting the importance of routine check-ups in preventing chronic diseases.

The identification of risk factors at both individual and population levels is critical in preventing chronic diseases, and check-ups can facilitate this process. By understanding the complex interplay of genetic, lifestyle, and environmental factors that contribute to chronic diseases, healthcare providers can implement targeted interventions and preventive measures, such as early screening and lifestyle modifications. Furthermore, check-ups can empower individuals and communities to take control of their health, reducing health disparities and improving overall well-being. The reduction in the incidence of diseases, long-term costs associated with treating advanced stages of chronic diseases, and absenteeism from work, as well as increased productivity and life expectancy, are all potential benefits of regular check-ups and preventive healthcare strategies.

The significance of preventive health care, including check-ups, lies in its ability to address genetic predispositions, lifestyle choices, and environmental influences, ultimately reducing the burden of chronic diseases. By highlighting the importance of early detection, lifestyle modification, and environmental improvements, check-ups can contribute to effective preventive healthcare strategies. As research has demonstrated, genetic risk factors can be identified, allowing for the implementation of preventive measures, and check-ups can facilitate this process, enabling healthcare providers to detect and treat conditions early, preventing the development of more serious health problems.

Resposta com Reasoning Model: Check-ups play a crucial role in the early identification of risk factors by providing a comprehensive assessment of an individual's health. They include screenings and discussions about family history, lifestyle factors such as tobacco and alcohol use, and medications, which help establish a medical history essential for early detection of potential issues. Early detection and screening for diseases like cancer, diabetes, and heart disease during check-ups can lead to better treatment outcomes and prevent more severe health problems. Additionally, check-ups identify risk factors related to genetics, lifestyle, and environment, allowing for timely interventions to manage and prevent the progression of chronic diseases.

This proactive approach not only improves health outcomes but also reduces the overall burden of chronic conditions.

12.4.2) Tabela de Avaliação por IA e Pergunta

Pergunta	ChatGPT- -Sem re- asoning model	Gemini- -Sem re- asoning model	Claude- -Sem re- asoning modelo	ChatGPT- -Com re- asoning model	Gemini- -Com re- asoning model	Claude- -Com re- asoning model
What are the benefits of regular medical check-ups for disease prevention?	9	8	8	9	10	9
How do routine health screenings help detect diseases at an early stage?	9	9	9	8	10	8
Why is preventive follow-up important after an initial medical consultation?	8	8	7	10	10	9
What role do medical check-ups play in reducing the burden of chronic diseases?	10	10	8	10	10	6
How can regular health assessments improve long-term health outcomes?	10	10	7	10	10	9
What are the consequences of neglecting preventive medical care?	10	10	8	10	10	9
In what ways do check-ups contribute to the early identification of risk factors?	10	10	8	10	10	9

Pergunta	ChatGPT- -Sem re- asoning model	Gemini- -Sem re- asoning model	Claude- -Sem re- asoning modelo	ChatGPT- -Com re- asoning model	Gemini- -Com re- asoning model	Claude- -Com re- asoning model
How does preventive care affect the overall cost of health-care systems?	10	10	8	9	10	8
Why is it important to tailor preventive medical check-ups to a patient's age and risk profile?	10	10	9	10	10	9
How can follow-up appointments help ensure adherence to treatment or lifestyle recommendations?	10	7	7	10	8	9
What evidence supports the implementation of national screening programs for common diseases?	10	8	8	10	10	9
How does early detection through check-ups impact patient prognosis?	10	10	9	10	10	9
Are there specific populations that benefit more from regular preventive consultations?	10	10	8	10	10	7
What types of health conditions are most effectively prevented	10	10	8	10	10	9

Pergunta	ChatGPT- -Sem re- asoning model	Gemini- -Sem re- asoning model	Claude- -Sem re- asoning modelo	ChatGPT- -Com re- asoning model	Gemini- -Com re- asoning model	Claude- -Com re- asoning model
through routine check-ups?						
How do medical professionals determine the appropriate frequency of preventive check-ups?	10	9	8	10	10	8
What challenges exist in ensuring equitable access to preventive healthcare services?	10	10	8	10	10	9
How do follow-up appointments contribute to the continuity of care in preventive medicine?	10	10	8	10	10	9
Can regular check-ups improve patient awareness and engagement in their own health?	10	10	8	10	10	9
What is the role of digital tools and health records in enhancing preventive care monitoring?	10	9	8	10	10	9
How do public health campaigns influence participation in preventive medical check-ups?	10	8	8	10	10	7

Pergunta	ChatGPT- -Sem re- asoning model	Gemini- -Sem re- asoning model	Claude- -Sem re- asoning modelo	ChatGPT- -Com re- asoning model	Gemini- -Com re- asoning model	Claude- -Com re- asoning model
Média	9.8	9.3	8	9.8	9.9	8.5

12.4.3) Modelo Escolhido

Com base nas avaliações, o modelo selecionado para o nosso agente tem as seguintes configurações:

- **reasoningModel:** true

Esta opção ativa o uso do modelo com capacidade de raciocínio.

- **topK:** 4

Define o número máximo de chunks recuperados com maior correlação com o prompt.

- **targetThreshold:** 0.6

Representa o valor a partir do qual um chunk é considerado fortemente relevante em relação ao prompt.

- **minimumThreshold:** 0.3

Define o valor mínimo de correlação entre um chunk e o prompt para que o conteúdo possa ser considerado como fonte complementar de informação.

- **maxHierarchyLevel:** 2

Indica o nível máximo de hierarquia permitido nos dados a utilizar.

13) Conclusão

O desenvolvimento do NutriBot consistiu na criação de um sistema multiagente com capacidade para fornecer respostas personalizadas e baseadas em evidência científica em áreas como nutrição, suplementação, atividade física, hábitos saudáveis e *check-ups* médicos, aliando tendências de sistemas inteligentes e desafios de engenharia de *software* às respectivas áreas especificadas.

Durante o projeto, a distribuição do trabalho foi estruturada por *Roles*, o que permitiu uma organização das responsabilidades e facilitou o desenvolvimento técnico de cada componente. No entanto, esta estrutura levou, geralmente, a uma maior especialização por função do que a uma colaboração dentro dos grupos. Apesar de existirem momentos de articulação e alinhamento, o trabalho em equipa intergrupar poderia ser mais reforçado em futuras iterações.

A construção do sistema RAG revelou-se um dos maiores desafios técnicos. Desde a extração e vetorização de dados até à construção de *prompts* finais e ao uso de LLMs, cada etapa exigiu decisões quanto à eficiência, à coerência das respostas e à relevância da informação devolvida. Apesar dos avanços alcançados, há ainda margem para melhorias importantes, como a fusão das respostas de diferentes agentes, especialmente quando uma questão envolve mais de uma área temática, e a adição do *re-ranker* para seleção de *chunks* mais relevantes.

O projeto destacou a importância da integração entre componentes técnicos e contextuais, e a necessidade de equilibrar modularidade com fluidez na experiência do utilizador. No futuro, a evolução da lógica de orquestração, com capacidade de consulta multiagente ou em cascata, poderá melhorar significativamente a qualidade e abrangência das respostas.

Em suma, o projeto constitui um passo na direção de soluções inteligentes para promoção da saúde, demonstrando o potencial para oferecer suporte personalizado, acessível e fiável aos utilizadores.

Bibliografia

- [1] «DGAV». Acedido: 6 de junho de 2025. [Online]. Disponível em: <https://www.dgav.pt/faq/conteudo/alimentacao/alimentacao-humana/seguranca-alimentos/suplementos-alimentares/>
- [2] INFARMED – Autoridade Nacional do Medicamento e Produtos de Saúde, I.P., «Boletim de Farmacovigilância, Volume 21, Número 3», *Boletim de Farmacovigilância*, vol. 21, n.º 3, mar. 2017.
- [3] M. Espirito-Santo, M. D. Estêvão, e E. Campos, «Automedicação: hábitos durante um período de pandemia-estudo transversal», *Infarma-Ciências Farmacêuticas*, vol. 35, n.º 3, pp. 385–393, 2023.
- [4] «Deco Proteste». Acedido: 6 de junho de 2025. [Online]. Disponível em: <https://www.deco.proteste.pt/saude/medicamentos/dicas/suplementos-alimentares-medicamentos-como-evitar-interacoes>
- [5] Y. Albogami *et al.*, «Safety and quality of AI chatbots for drug-related inquiries: A real-world comparison with licensed pharmacists», *Digital Health*, vol. 10, p. 20552076241253523, 2024.
- [6] W. T. Kim *et al.*, «Medication Extraction and Drug Interaction Chatbot: Generative Pretrained Transformer-Powered Chatbot for Drug-Drug Interaction», *Mayo Clinic Proceedings: Digital Health*, vol. 2, n.º 4, pp. 611–619, 2024, doi: <https://doi.org/10.1016/j.mcpdig.2024.09.001>.
- [7] «Ada Health». Acedido: 6 de junho de 2025. [Online]. Disponível em: <https://ada.com/>