



DISCERNMENT & DILIGENCE

Avaliação Crítica e Responsabilidade

⚡ Mostly Harmless? Será? ⚡



Módulo 3 — Discernment + Diligence

O Ciclo Completo: Description + Discernment

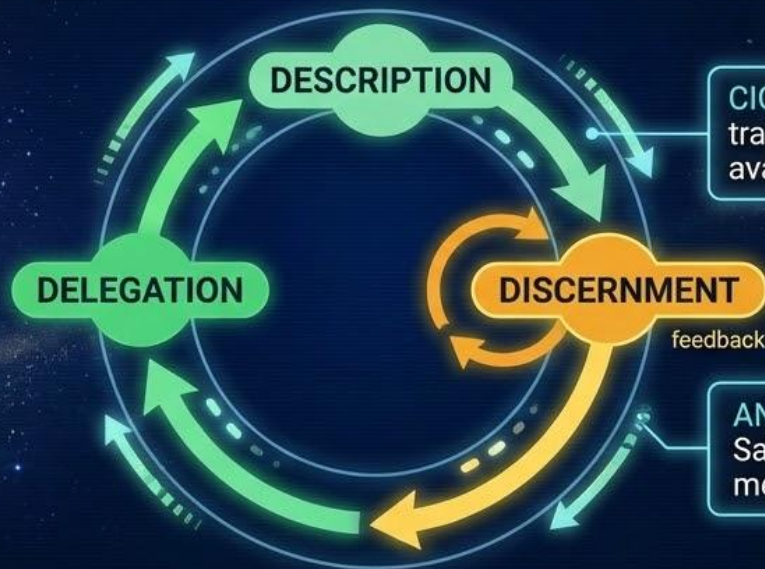
✓ Módulo 1: Decidimos
O QUE fazer com IA



✓ Módulo 2: Aprendemos
COMO comunicar



🎯 Módulo 3: AVALIAMOS o
que recebemos + USAMOS
com responsabilidade



CICLO: Description e Discernment
trabalham juntos: você pede,
avalia, ajusta, repete

feedback para Description

ANALOGIA: Saber pedir é metade.
Saber avaliar a resposta é a outra
metade.

75% Completo



Descrição + Discernimento = **Não se perder no espaço**



Terceiro D: DISCERNMENT

Seu sistema de controle de qualidade



DISCERNMENT

O que recebi está correto e é útil?

“O Guia nem sempre está certo. A IA também não.”

⚡ Verificação de fatos: modo ativado ⚡

O Que É Discernment?



Discernment é sua capacidade de avaliar criticamente o que a IA produz, como ela produz, e como ela se comporta.



É seu **SISTEMA DE CONTROLE DE QUALIDADE** para colaboração com IA

POR QUE É ESSENCIAL:

"Confie, mas verifique. Especialmente se a fonte for um livro intergaláctico."

- 🔍 Identificar quando saídas são valiosas vs. problemáticas
- ⚖️ Reconhecer forças e limitações
- ✅ Determinar quando está pronto para uso ou precisa de mais trabalho



REQUISITOS: Discernment requer **EXPERTISE DE DOMÍNIO** + compreensão de como IA funciona



1. DISCERNIMENTO DO PRODUTO

Avaliar a qualidade do que a IA realmente PRODUZ

PERGUNTAS A FAZER:

	Aspecto	Pergunta		
1.	Precisão	Isso é factualmente preciso?	<input checked="" type="checkbox"/>	✓
2.	Adequação	É apropriado para meu público e propósito?	<input checked="" type="checkbox"/>	✓
3.	Coerência	É coerente e bem estruturado?	<input checked="" type="checkbox"/>	✓
4.	Conformidade	Atende aos meus requisitos?	<input checked="" type="checkbox"/>	✓
5.	Valor	Adiciona valor ou resolve o problema?	<input checked="" type="checkbox"/>	✓

VERIFICADO



STATUS:

APROVADO/REJEITADO



Respostas podem parecer plausíveis e estar completamente erradas








Lembre-se: A toalha não é opcional. A verificação é.

2. DISCERNIMENTO DO PROCESSO

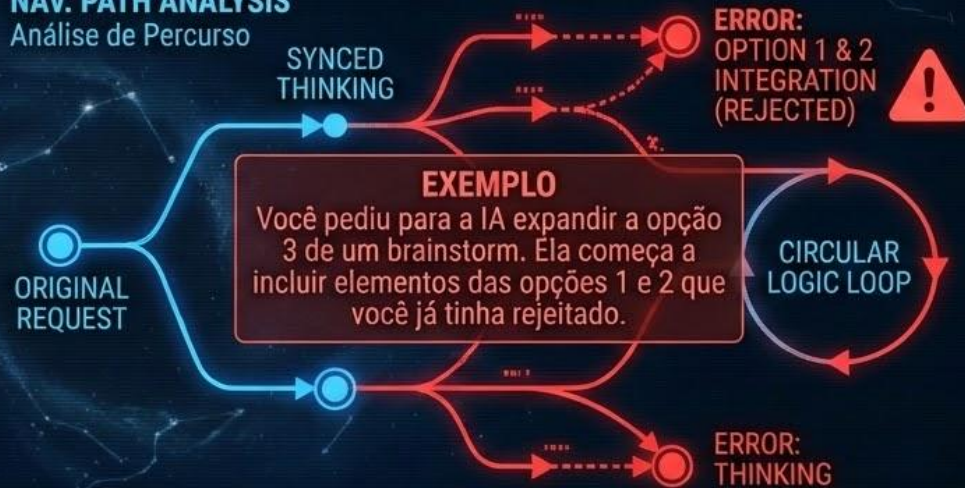
 **DEFINIÇÃO:** "Avaliar COMO a IA chegou na resposta, não só O QUE ela disse"

PROBLEMAS A IDENTIFICAR

- ****Erros lógicos**** no raciocínio 
- ****Lapsos de atenção**** – ignorou informações importantes 
- ****Passos inadequados**** – abordagem errada para o problema 
- ****Fixação**** – presa em uma interpretação sem considerar alternativas 
- ****Raciocínio circular**** – conclusão justifica premissa que justifica conclusão 

NAV. PATH ANALYSIS

Análise de Percurso



IMPORTÂNCIA: "Garante que você e a IA estejam 'pensando em sincronia'"

Rota recalculando... mas para o destino certo?
VEJA O CAMINHO, NÃO SÓ O DESTINO. DON'T PANIC!

3. Discernimento do Desempenho

DEFINIÇÃO: “Avaliar como a IA está se COMPORTANDO durante a interação”



DIFERENÇA SUTIL:

- **Processo** = o trabalho que a IA está fazendo
- **Desempenho** = quão bem ela está interagindo COM VOCÊ enquanto faz

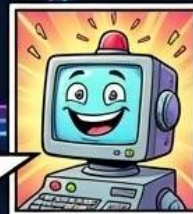
PERGUNTAS A FAZER

	Comunicação	Há uma forma melhor dela se comunicar?		Ajustar clareza?
	Utilidade	Está fornecendo informações de forma útil?		Mais prático?
	Responsividade	Responde bem ao seu feedback?		Ouvir melhor?
	Eficiência	A interação é eficiente ou desnecessariamente complexa?		Simplificar?
	Calibração	Está detalhada demais ou resumida demais?		Equilibrar?

➔ **AÇÃO:** “Se o desempenho não está bom, use *Description* para ajustar\”

Marvin: comunicação excessivamente deprimida.

Eddie: otimismo exagerado. Ajuste conforme necessário.



Os 3 Tipos de Discernment

 **Produto** 



Foco	Avaliar saídas
Pergunta Central	"Isso está correto e é útil?"

 **CORRETO & ÚTIL**


 **Processo** 

Foco	Avaliar raciocínio
Pergunta Central	"A IA chegou aqui de forma lógica?"

 **RACIOCÍNIO LÓGICO** 

 **Desempenho** 

Foco	Avaliar interação
Pergunta Central	"A IA está colaborando bem comigo?"

 **COLABORAÇÃO EFETIVA**



Três tipos de verificação: porque dois não eram suficientes.
SEU PAINEL DE CONTROLE INTEGRADO. DON'T PANIC!

Red Flags: Quando Desconfiar

DANGER - ALERTA

DANGER - ALERTA

DANGER - ALERTA

DANGER - ALERTA

DANGER - ALERTA



Confiança excessiva em fatos específicos

Datas, números, citações exatas — alta chance de alucinação



Referências bibliográficas detalhadas

Autor, ano, revista, páginas — frequentemente inventados



Afirmações sobre eventos recentes

Data de corte do conhecimento — pode não saber



Unanimidade ou certeza absoluta

"Todos concordam que..." — simplificação perigosa



Resposta que parece "perfeita demais"

Muito alinhada com o que você queria ouvir — viés de confirmação



REGRA DE OURO

Se é importante, verifique. Se é crítico, verifique duas vezes.

A Vagon poetry não é a única coisa perigosa no universo. Mantenha um olho vivo!



DILIGENCE

Responsabilidade pelo que você faz com IA

Central Question

Estou usando IA de forma responsável?

— A Hitchhiker's Commitment



Dirigir um carro não é só ir do ponto A ao B.
É seguir regras, considerar outros na estrada, assumir responsabilidade.

***"Com grandes poderes (de IA)
vêm grandes responsabilidades."***

— Tio Ben, provavelmente

O Que É Diligence?

DEFINITION: Diligence é assumir responsabilidade por suas interações com IA — garantindo uso rigoroso, transparente e responsável.

QUESTÕES CENTRAIS



Quais são as implicações de trabalhar com esta IA?



Quem pode ser afetado pelo que é criado?



Quem tem acesso aos dados que estou compartilhando?



Minha interação está alinhada com meus valores e padrões éticos?

OS TRÊS COMPONENTES



1. Diligência na Criação
escolha de sistemas e proteção de dados



2. Diligência na Transparência
honestidade sobre uso de IA



3. Diligência na Implantação
responsabilidade pelo resultado final



Mais importante que saber onde sua toalha está



1. Diligência na Criação

Ser crítico e intencional sobre **QUAIS** sistemas de IA você usa e **COMO** usa

SAFETY PROTOCOL: DILIGENCE PRE-FLIGHT CHECK

PERGUNTAS A FAZER: CHECKLIST PRÉ-VOO



Sobre o Sistema

Como foi treinado? Que dados foram usados?



Verificado



Sobre Seus Dados

Quem é dono do que eu inserir? Quem pode acessar?



Verificado



Sobre Privacidade

Estou protegendo a privacidade de outros?



Verificado



Sobre Valores

Isso se alinha com meus valores e políticas da organização?



Verificado

EXEMPLO PRÁTICO:



Antes de compartilhar dados sensíveis de clientes com uma IA, verifique: a empresa permite isso? O serviço tem políticas adequadas de proteção?

Não compartilhe segredos com estranhos. Nem com IAs desconhecidas.

2. Diligência na Transparência

DEFINIÇÃO: Ser honesto sobre o papel da IA no seu trabalho com todos que precisam saber

PERGUNTAS-GUIA



Quem precisa saber sobre o papel da IA?



Como e quando devo comunicar isso?







Que nível de detalhe faz sentido?



POR QUE IMPORTA: Não é só seguir regras — é **MANTER CONFIANÇA** **CONFIANÇA E RESPEITO** em relacionamentos

CONTEXTOS DIFERENTES, EXPECTATIVAS DIFERENTES

Contexto	Expectativa típica
 Acadêmico	Declaração explícita exigida
 Profissional	Varia por organização/indústria
 Criativo	Normas ainda em formação
 Pessoal	A seu critério

Honestidade: mais rara que água no deserto de Kakrafoon

HITCHHIKER'S GUIDE Inspired

3. Diligência na Implantação

Assumir **RESPONSABILIDADE FINAL** pelo que você compartilha com o mundo



PRINCÍPIO CENTRAL

A IA pode cometer erros. Quando você publica algo, VOCÊ é responsável — não a IA

ISSO SIGNIFICA:

- ✓ Verificar fatos antes de publicar
- ✓ Checar vieses e imprecisões
- ✓ Garantir precisão e adequação
- ✓ Confirmar direitos de uso (imagens, textos)
- ✓ Aplicar os mesmos padrões que se você tivesse feito sozinho

EXEMPLO

Um jornalista usa IA para redigir um artigo. Antes de publicar, ele verifica cada fato e fonte — os mesmos padrões de sempre. A IA ajudou, mas ele assina e responde pelo conteúdo.






Sua Assin

Sua Assinatura Aqui

Não existe "a IA errou". Existe "você publicou sem verificar".

Os 3 Componentes de Diligence

TABELA SÍNTESE

Componente	Foco	Responsabilidade
 Criação	Escolha de sistemas	Ser intencional sobre quais IAs usar e como
 Transparência	Divulgação	Ser honesto sobre o papel da IA
 Implantação	Responsabilização	Assumir responsabilidade pelo resultado final

MENSAGEM CENTRAL

Diligence garante que sua colaboração com IA seja não apenas **EFICAZ** e **EFICIENTE**, mas também **ÉTICA** e **SEGURA**



Selo de Mochileiro Responsável: nível Polegar de Ouro



Declaração de Diligência





CONCEITO: Um reconhecimento transparente do papel da IA no seu trabalho

EXEMPLO DE ESTRUTURA:



Na criação deste [documento/projeto], colaborei com [IA] para auxiliar em [tarefas específicas]. Todo conteúdo gerado passou por revisão e validação. O resultado final reflete meu entendimento e expertise. Embora a assistência de IA tenha sido instrumental, mantenho total responsabilidade pelo conteúdo.

QUANDO USAR:

-  Trabalhos acadêmicos (frequentemente obrigatório)
-  Projetos profissionais (quando apropriado)
-  Publicações (transparência como valor)
-  Este próprio curso (veja a página "Sobre")



Declaração de Diligência: mais útil que 'Mostly Harmless'

O Ciclo Completo em Ação




Na prática, você passa por este ciclo várias vezes até chegar no resultado ideal

Mais eficiente que o Drive de Improbabilidade. Menos aleatório também.




Exercício: Identifique os Problemas

Analise esta resposta de IA e identifique potenciais problemas

 **under scrutiny** 

De acordo com o estudo de Martinez et al. (2023) publicado no Journal of AI Ethics, 87% dos profissionais que usam IA regularmente reportam aumento de 40% na produtividade. O estudo, conduzido com 5.000 participantes em 12 países, é considerado definitivo na área. Todos os especialistas concordam que a IA é essencial para o trabalho moderno.

PERGUNTAS:

1.  Que red flags você identifica?
2.  O que você verificaria antes de usar?
3.  O que parece confiável vs. suspeito?

 Tempo: 5 minutos

Achou fácil? As alucinações bem feitas são mais sutis.

Terceiro e Quarto D's Completos!

95% COMPLETED

ALMOST THERE!



✓ Hoje aprendemos:

- **Discernment:** Avaliar Produto, Processo e Desempenho
- Red flags para identificar problemas
- **Diligence:** Criação, Transparência e Implantação
- Declaração de Diligência como prática

📌 Próximo módulo (último!):

- Síntese de tudo
- Plano pessoal de fluência
- Visão do futuro

SYNTHESIS & APPLICATION



Vocês agora sabem: O QUE, COMO, AVALIAR e SER RESPONSÁVEL. Falta só juntar tudo.

Você agora é oficialmente um mochileiro da IA. Só falta a toalha.