



Fillipe Guerra &lt;fillipe.backup@gmail.com&gt;

## EAAS Whitepaper 02

1 mensagem

Fillipe Guerra &lt;fillipe.backup@gmail.com&gt;

25 de outubro de 2025 às 23:48

Para: Fillipe Augusto Gomes Guerra &lt;fillipe182@hotmail.com&gt;, Fillipe Guerra &lt;fillipe.backup@gmail.com&gt;

# Capítulo 1 — Prólogo e Filosofia da Plataforma

Fillipe Guerra - 2025

## 1.1 Origem e Propósito

EAAS nasce da observação de um fenômeno recorrente na história corporativa: à medida que as empresas se digitalizaram, multiplicaram-se também os seus sistemas, bases de dados e silos de decisão.

CRM, ERP, BI, marketplace e atendimento cresceram como ilhas de informação. O resultado foi fragmentação: decisões lentas, custos altos e inteligência dispersa.

EAAS propõe o oposto: **um único organismo digital**, capaz de **perceber, decidir e agir** sobre cada operação empresarial — seja ela financeira, logística, comercial ou relacional — com a mesma fluidez com que um cérebro coordena o corpo.

## 1.2 O Princípio da Integração Viva

O sistema EAAS não é apenas uma coleção de módulos.

Ele é concebido como um **ecossistema adaptativo**:

- Cada módulo (CRM, ERP, Marketplace, Omnichat) é uma função vital.
- O **Núcleo Cognitivo da IA** atua como o sistema nervoso, aprendendo com cada sinal de uso.
- A infraestrutura multi-tenant é o esqueleto que sustenta milhares de empresas isoladas, mas metabolicamente semelhantes.

O lema da EAAS é:

*“Tudo como serviço. Inteligência como substância.”*

## 1.3 Visão Humanista e Ética

A EAAS adota o paradigma da **inteligência ampliada**, não da substituição.

A IA deve:

1. **Aumentar** a capacidade humana de decidir.
2. **Preservar** o contexto ético, jurídico e social.
3. **Gerar valor sustentável**, não apenas eficiência.

Matematicamente, esse equilíbrio se modela pela função de valor tripla:

$$U = \alpha U_{\text{econômico}} + \beta U_{\text{humano}} - \gamma U_{\text{ético\_penalty}}$$

com pesos  $\alpha, \beta, \gamma$  ajustáveis por política corporativa.

O núcleo de IA buscará maximizar  $U$  sob restrições de conformidade e bem-estar.

## 1.4 O Paradigma Neural-Empresarial

A arquitetura da EAAS introduz o conceito de **Plataforma Neural de Negócios**, em que:

- **dados** são sinapses,
- **serviços** são neurônios especializados,
- **a IA** é o córtex preditivo e decisório,
- **as empresas-tenants** são organismos distintos, todos alimentados pela mesma infraestrutura cognitiva.

A cada interação — um pagamento, uma reserva, uma mensagem — a rede aprende, ajusta pesos e refina políticas. Não há separação entre “sistema de informação” e “inteligência”: há apenas graus de consciência digital.

1.5 Metas de Design

1. **Universalidade:** servir de base a qualquer ramo de atividade.
2. **Modularidade:** cada função é independente, mas interoperável via API Gateway.
3. **Auto-aprendizagem:** a IA adapta fluxos e recomendações em tempo real.
4. **Explicabilidade:** toda decisão é rastreável, com causalidade formal.
5. **Soberania de dados:** multi-tenant isolado, privacidade diferencial e compliance LGPD.
6. **Elegância:** simplicidade visual e coerência matemática; o sistema deve “respirar”.

1.6 Estrutura Deste Documento

Este whitebook é dividido em quatro grandes domínios:

Domínio	Conteúdo
Arquitetura Funcional	Descrição de todos os módulos (Marketplace, CRM, ERP, Omnichat, Pagamentos, Calendário, Administração Multiempresa).
Arquitetura Cognitiva	IA autônoma, neuro-simbólica e ética; aprendizado federado; estabilidade e autocorreção.
Matemática e Modelagem	Fundamentos formais: otimização, reforço, Lyapunov, lógica LTL+D, prova de convexidade e coerência ética.
Governança e Implementação Conceitual	Padrões de segurança, observabilidade, métricas, roadmap e princípios de evolução.

Capítulo 2 — Visão Geral e Princípios de Design da Arquitetura EAAS

2.1 O Conceito de Ecossistema Modular

EAAS é estruturada como um **sistema de sistemas**. Cada empresa (“tenant”) possui um conjunto próprio de módulos, mas todos partilham um mesmo núcleo de padrões:

- **Camada de Experiência:** interfaces humanas (Marketplace, painéis, Omnichat).
- **Camada de Aplicação:** CRM, ERP, Financeiro, Calendário.
- **Camada de Serviços Cognitivos:** motor de IA e regras simbólicas.
- **Camada de Dados:** bancos relacionais, vetoriais e grafo semântico.
- **Camada de Infraestrutura:** Kubernetes, API Gateway, autenticação e observabilidade.

Matematicamente, o sistema é um grafo dirigido

$G=(V,E)$

onde cada vértice  $v_i$  representa um módulo e cada aresta  $e_{ij}$  um fluxo de dados ou evento. A função de coerência global é:

$$\Phi(G) = \sum_{(i,j) \in E} w_{ij} C_{ij}$$

onde  $C_{ij}$  mede a compatibilidade semântica entre módulos  $i$  e  $j$ .

A meta de design é maximizar  $\Phi(G)$  — integração — sob restrições de isolamento e segurança.

## 2.2 Princípios Estruturantes

1. **Isomorfismo Operacional:** todos os tenants compartilham a mesma topologia lógica.
2. **Separação de Preocupações:** funções de negócio, dados e IA independentes, comunicando-se via contratos formais.
3. **Elasticidade:** recursos de computação adaptam-se dinamicamente à carga.
4. **Auditabilidade:** cada evento é rastreável; nenhuma decisão é opaca.
5. **Sustentabilidade:** eficiência energética e econômica como variáveis do modelo.

## 2.3 Multi-Tenancy e Isolamento

Cada empresa  $T_k$  possui seu subconjunto de dados  $D_k$  e políticas  $P_k$ .  
Formalmente:

$$D_i \cap D_j = \emptyset, \forall i \neq j$$

A agregação cognitiva ocorre via aprendizado federado:

$$\theta_{\text{global}} = \sum_k \frac{n_k}{n} \theta_k$$

preservando a privacidade dos dados locais.

## 2.4 Fluxos de Informação

Um evento de negócio segue o ciclo:

1. **Input humano ou automatizado** (ex.: mensagem ou compra).
2. **Roteamento semântico** (Omnichat  $\rightarrow$  IA  $\rightarrow$  módulo adequado).
3. **Processamento** (ERP, CRM, Financeiro).
4. **Atualização de conhecimento** (IA registra padrões).
5. **Feedback** (dashboards, relatórios, previsões).

O tempo médio do ciclo  $t_c$  é otimizado por:

$$\text{config}_{\text{mintc}} = f(\text{latência rede}, \text{complexidade workflow}, n_{\text{chamadas}})$$

## 2.5 Modelo de Comunicação

A comunicação entre módulos usa **mensageria assíncrona**.

Cada mensagem  $m$  contém: identificador, tenant, timestamp, assinatura.

A taxa de perda  $p_l$  é monitorada; a confiabilidade do sistema é

$$R(t) = e^{-\lambda t}$$

onde  $\lambda$  é a taxa de falhas não recuperadas.

Para SLO 99.9 %, exige-se  $\lambda < 10^{-6}$  falhas/s.

## 2.6 Governança de Dados e Ética

O grafo de dados é anotado com *tags* de privacidade:

$$\text{tag}(x) \in \{\text{PII}, \text{Financeiro}, \text{Público}\}$$

e as políticas de acesso são funções booleanas  $\text{access}(u, x, t)$  definidas em Lógica Deontica Temporal:

$$\Box(\neg \text{consent}(u, t) \rightarrow \neg \text{collect}(x))$$

## 2.7 Interoperabilidade e API Gateway

Todas as interações externas passam pelo Gateway:

- autenticação OAuth2/JWT;
- limitação de taxa  $r \leq r_{max}$ ;
- validação de *schema*;
- registro de métricas OpenTelemetry.

O Gateway implementa a regra de integridade:

$\forall request, valid(request) \Rightarrow consistent(response)$

2.8 Resiliência e Continuidade

A estabilidade global do sistema é garantida por um critério de Lyapunov discreto:

$V(t+1) - V(t) \leq -\epsilon \|x_t\|^2$

onde  $V(t)$  é a energia de falhas acumulada.  
Quando  $\epsilon > 0$ , o sistema converge para estabilidade assintótica, mesmo sob perturbações.

2.9 Resumo Filosófico do Design

A plataforma é desenhada segundo o princípio da **simplicidade coerente**:  
quanto mais homogêneas as interfaces e mais formalizadas as dependências, mais fácil é escalar, corrigir e aprender.

Capítulo 3 — Arquitetura Macro do Ecossistema EAAS

3.1 Estrutura em Camadas

A plataforma é organizada em cinco camadas funcionais interligadas, cada uma com sua função precisa:

Camada	Responsabilidade	Componentes principais
Frontend	Interação com usuários finais e operadores	Portais Web, Aplicativos Móveis, Widgets do Marketplace
Aplicação (Backend)	Regras de negócio e processos	CRM, ERP, Financeiro, Calendário, Admin
Camada Cognitiva (IA)	Percepção, planejamento e decisão	Planner, Críticos, Memórias, RAG
Camada de Dados	Persistência e coerência semântica	PostgreSQL, Redis, Vetores, Grafos
Infraestrutura e DevOps	Escalabilidade, segurança e observabilidade	Kubernetes, API Gateway, Vault, OpenTelemetry

O fluxo básico é:

FrontendEventosBackendMensageriaAAC\co~esDados/ERP/CRM

e a resposta retroalimenta as camadas anteriores, gerando aprendizado.

3.2 Frontend – Camada de Experiência

O Frontend segue princípios de usabilidade clara e modularidade visual:

- Portais administrativos para empresas;
- Marketplace aberto a clientes;
- Chat on-page integrado à IA e ao Omnichat;
- Responsividade total (móvel ↔ desktop).

Cada interação gera eventos com identidade de tenant e contexto semântico:

$E=\{\text{user\_id}, \text{tenant\_id}, \text{action}, \text{timestamp}, \text{payload}\}$

A métrica de sucesso de UX é definida como:

$SUX = \text{ac,ões iniciadas} \text{ac,ões concluídas} \times 1 + \text{latência média}$

---

### 3.3 Backend – Camada de Aplicação

É o coração transacional.

Cada microserviço segue contratos formais (API Schemas JSON) e publica eventos para a fila de mensageria.

Os módulos principais — CRM, ERP, Financeiro, Calendário — têm interfaces idênticas de autenticação e auditoria.

A consistência entre serviços é mantida pelo modelo de eventual consistency:

$P(\text{inconsistência em } t) = e^{-kt}$

onde  $k$  é a taxa de replicação.

---

### 3.4 Camada Cognitiva – IA Autônoma

O núcleo cognitivo atua como mecanismo de decisão.

É composto por:

- **Planner** – decompõe tarefas em etapas sequenciais.
- **Critic** – valida respostas (factualidade, ética, numérica).
- **Executor** – invoca ferramentas dos módulos de negócio.
- **Memory Layer** – mantém contexto de curto e longo prazo.
- **Knowledge Base** – vetores e grafos semânticos.

O aprendizado é um problema de otimização multivariável:

$\theta \min J(\theta) = E(x, y) [\ell(f(\theta(x), y))] + \lambda_1 R_{\text{stability}} + \lambda_2 R_{\text{ethic}}$

onde  $R_{\text{stability}}$  e  $R_{\text{ethic}}$  são regularizadores de estabilidade e ética.

A condição de estabilidade é garantida por um funcional de Lyapunov:

$V(t+1) - V(t) \leq -\epsilon \|g_t\|^2, \epsilon > 0$

---

### 3.5 Camada de Dados

Integra três tipos de armazenamento:

1. Relacional (SQL) para transações fortes.
2. Vetorial para busca semântica.
3. Grafo para relações entre entidades.

O grafo semântico é definido por  $(V, E)$  com peso  $w_{ij}$  representando relevância.

As consultas usam algoritmos de caminho mínimo e propagação de confiança:

$c_i = \sum_j w_{ij} c_j$

que resolve-se por iterações até convergir ( $\|c(t+1) - c(t)\| < \delta$ ).

---

### 3.6 Infraestrutura e DevOps

Conjunto de mecanismos para garantir disponibilidade e segurança:

- **Orquestração:** Kubernetes com Auto-Scaler.
- **CI/CD:** pipelines versionadas, rollback atômico.
- **Segurança:** Vault (KMS), TLS, RBAC.

- **Observabilidade:** OpenTelemetry → Prometheus → Grafana.
- **Resiliência:** backup incremental, testes de recuperação.

Métrica de eficiência energética:

$\eta = \text{energia consumida} / \text{operações úteis} [\text{ops/kWh}]$

Objetivo: maximizar  $\eta$  sem reduzir SLO de 99.9 %.

### 3.7 Relação entre Camadas

Cada camada expõe interfaces claras:

$l_i \circ l_{i+1} \approx l_d$

garantindo composicionalidade.

A incoerência média entre camadas:

$\Delta^- = N^{-1} \sum_i \|l_i(l_{i+1}(x)) - x\|$

deve permanecer menor que um limite  $\delta_{\text{design}}$ .

### 3.8 Resumo

A arquitetura EAAS é uma topologia viva: as camadas trocam informações como sistemas biológicos trocando sinais metabólicos. A IA atua como homeostato, mantendo equilíbrio entre demanda, recurso e aprendizado.

## Capítulo 4 — Marketplace Universal e CRM 360°

### 4.1 Marketplace Universal: a Vitrine Cognitiva

O Marketplace EAAS é uma estrutura neutra capaz de vender **produtos, serviços e experiências**. Cada empresa-tenant possui sua instância própria do catálogo, mas a lógica de negócio é compartilhada.

#### 4.1.1 Estrutura do Catálogo

O catálogo é uma tabela dinâmica com atributos mínimos:

Campo	Descrição
sku	Identificador único
type	Produto / Serviço / Experiência
title, description	Dados semânticos indexados
price, currency	Valores transacionais
addons	Complementos configuráveis
availability	Slots ou estoque vinculados ao Calendário
tenant_id	Escopo de empresa

A IA classifica cada item em um grafo de relação:

$P(i,j) = \sigma(w_i \cdot w_j)$

onde  $P(i,j)$  mede a probabilidade de dois itens serem complementares; isso permite vendas cruzadas automáticas.

#### 4.1.2 Carrinho e Checkout

O processo segue a sequência:

1. Cliente escolhe item.
2. IA recomenda adicionais (por similaridade vetorial).

3. Totais calculados com regras de impostos e moedas.
4. Pagamento via Stripe/Pix com split automático.
5. Ordem confirmada → ERP e Calendário são atualizados.

Função de conversão:

$$\text{Conv} = \text{Vtotal} / \text{VIA}$$

representa a fração de vendas assistidas pela IA.

#### 4.1.3 Reservas e Experiências

Serviços e experiências possuem ligação com slots de tempo e recursos.

Um slot é um vetor  $(r_i, t_i, d_i)$  = recurso, tempo, duração.

A IA resolve a alocação minimizando:

$$C = \alpha t + \beta d + \gamma(1 - u)$$

onde  $u$  é taxa de uso do recurso.

#### 4.1.4 Pós-venda e Fidelização

Após a entrega, a IA avalia sentimento das mensagens e atribui pontuação de satisfação.

Modelo simplificado:

$$S = \tanh(ws \cdot \text{ftexto})$$

$S > 0.5 \Rightarrow$  positivo.

## 4.2 CRM 360° — O Coração da Relação com o Cliente

O CRM integra todas as interações, histórico e dados de comportamento.

### 4.2.1 Modelo de Contato

Cada cliente é um nó no grafo social:

$$G_c = (V_c, E_c)$$

com vértices  $V_c$ =clientes e arestas  $E_c$ =interações.

A centralidade de engajamento é:

$$E_i = \sum_j w_{ij}$$

usada para priorizar follow-ups.

### 4.2.2 Funil e Automação

O funil de vendas é um processo de Markov:

$$p_{t+1} = P p_t$$

onde  $P$  é a matriz de transição entre etapas.

A IA otimiza  $P$  para maximizar a probabilidade de “ganho”.

Automação básica (exemplo conceitual):

Se cliente inativo > 30 dias → gerar oferta personalizada.

Formalmente:

$$A(c) = \{1, 0, \text{se } \Delta t > 30 \text{ caso contrário}\}$$

### 4.2.3 Insight e Predição

Comportamento é analisado via regressão logística:

$$P(\text{compra}) = \sigma(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)$$

A IA gera scores de propensão e sugestões de ação comercial.

#### 4.2.4 Integração com Omnichat

Cada mensagem de chat é anexada ao histórico do cliente.

O modelo de sentimento agrega informações textuais à base numérica do CRM.

### 4.3 Coerência entre Marketplace e CRM

Os dois módulos compartilham as mesmas entidades: cliente, pedido, pagamento.

A função de consistência:

$$\Delta = \|D_{\text{market}} - D_{\text{crm}}\|$$

deve ser mínima; auditoria automática corrige desvios.

### 4.4 Interação com IA

A IA atua em quatro níveis:

1. Recomendações de venda e upsell.
2. Previsão de comportamento de clientes.
3. Detecção de anomalias em pedidos.
4. Geração de respostas no Omnichat.

A qualidade do modelo é avaliada por:

$$F1 = \text{precisão} + \text{recall} / 2$$

objetivo  $\geq 0.9$  para decisões automáticas.

### 4.5 Síntese

O Marketplace e o CRM formam o “lado de fora e dentro” da empresa.

O primeiro gera receita; o segundo mantém relacionamento.

Unidos pelo mesmo núcleo de IA, eles transformam transações em aprendizado contínuo.

## Capítulo 5 — ERP Completo e Gestão Financeira

### 5.1 Visão Geral do ERP EAAS

O ERP da EAAS é o **núcleo operacional contábil-logístico**. Integra: **Financeiro (AP/AR/Fluxo de Caixa/Dupla-Entrada)**, **Estoque & Logística**, **Contábil & Fiscal**, **RH & Folha**, e **BI**.

Seu desenho é **multi-tenant** e **event-driven**: cada lançamento contábil, cada entrada/saída de estoque ou evento de folha é **um fato auditável**, com **idempotência**.

### 5.2 Financeiro: Contas a Pagar/Receber, Caixa e Conciliação

#### 5.2.1 Dupla-Entrada (Double-Entry Ledger)

Cada evento financeiro gera lançamentos que mantêm a identidade

$$\sum \text{Débitos} = \sum \text{Créditos}$$

Seja um pedido de valor V (líquido) e imposto T. No recebimento:



- **Débito:** Caixa +(V+T)
- **Crédito:** Receita +V, Impostos a Recolher +T

A verificação de consistência do livro razão:

$$\forall j, i \sum L_{ij} \text{debit} - i \sum L_{ij} \text{credit} = 0,$$

garante **conservação contábil**.

### 5.2.2 Contas a Receber (AR) e a Pagar (AP)

Registra-se faturas com vencimento dv. O **aging** (escalonamento por atraso) calcula buckets  $B_k$  (ex.: 0–30, 31–60 dias...).

Saldo por bucket:

$$\text{Saldo}(B_k) = f \in B_k \sum \text{valor\_aberto}(f).$$

Essa decomposição alimenta previsão de caixa e políticas de cobrança.

### 5.2.3 Fluxo de Caixa e Projeções

Projeção de caixa no horizonte H:

$$C^t + h = C^t + i \in \text{entradas} \sum E[E_i, t+h] - j \in \text{saídas} \sum E[S_j, t+h],$$

onde as expectativas podem usar modelos de **sazonalidade** (ex.: Holt-Winters) ou **processos ARIMA**, e ajuste por **probabilidade de inadimplência**  $p_{\text{default}}$ .

## 5.3 Estoque & Logística

### 5.3.1 Saldos e Valoração

Cada SKU possui movimentações {in,out,reserve,release}.

Saldo:

$$Q_t = Q_{t-1} + \sum \text{in} - \sum \text{out} - \sum \text{reserve} + \sum \text{release}.$$

Valoração por custo médio:

$$c^t = Q_t c^{t-1} - 1 Q_{t-1} + \sum (\text{in} \cdot c_{\text{in}}),$$

ou FIFO/LIFO conforme política.

### 5.3.2 Roteirização e SLA

Para entregas/experiências, a logística otimiza tempo/distância:

$$\min(i, j) \sum d_{ij} \cdot x_{ij} s.a. j \sum x_{ij} = 1, i \sum x_{ij} = 1,$$

(atribuição/húngaro) ou variantes de VRP.

SLA é controlado por janelas de tempo; violações disparam **multas/sinalizadores**.

## 5.4 Contábil & Fiscal

Plano de contas hierárquico; livros (Diário, Razão); **balancetes**.

**Impostos:** motores de **regras por jurisdição** (alíquotas, substituição tributária).

Verificação de integridade: somatórios por natureza **batem** com o razão.

## 5.5 RH & Folha (Básico)

Cálculo de salários, comissões, benefícios; export para sistemas locais.

Política de **segregação de acesso** a dados sensíveis (salários/PII).

Métricas de custo de pessoal por centro de custo.

## 5.6 BI & Dashboards

Camada analítica (OLAP) com medidas: **Receita, Margem, Ticket Médio, Giro de Estoque, DSO (Days Sales Outstanding), SLA.**

Acurácia do BI:

$$\epsilon_{BI} = \frac{\|Rel_{conta\_bil} - Rel_{BI} - Rel_{conta\_bil}\|}{\|Rel_{conta\_bil}\|} \leq \delta,$$

com  $\delta$  estipulado (ex.: 1%).

## 5.7 Conciliação com Pagamentos (Stripe/Pix/Boletos)

Conciliação via **webhooks idempotentes** e **referências cruzadas**.

Erro de conciliação:

$$\Delta_{conc} = \sum \text{stripe\_settled} - \sum \text{ERP\_recebido},$$

almejando  $\Delta_{conc} \rightarrow 0$  diariamente.

# Capítulo 6 — Pagamentos, Split e Payouts (Stripe Connect) + Conformidade

## 6.1 Fluxo Financeiro no Marketplace

Quando a venda envolve múltiplos parceiros, o **split** define a distribuição:

$$V_{total} = V_{plataforma} + p \sum V_{parceiro, p}.$$

**Stripe Connect** implementa as transferências (payouts) para contas conectadas, com **regras configuráveis** por SKU/loja/parceiro.

## 6.2 Política de Split e Comissões

Podemos definir pesos  $\omega_p$  por parceiro tal que

$$p \sum \omega_p = 1, V_{parceiro, p} = \omega_p \cdot V_{liquido}.$$

Custos e taxas (plataforma, gateway) compostos na **margem**:

$$\text{Margem} = V_{liquido} - (\text{taxas} + \text{custos diretos}).$$

## 6.3 Reembolsos, Disputas e Risco

Modelos de **anomaly detection** monitoram transações.

Penalidade por risco:

$$R_{fraude} = P(\text{chargeback}) \cdot \text{impacto\_econo}^{\text{mico}}.$$

Políticas exigem escalonamento humano acima de limiares.

## 6.4 Conformidade (PCI/LGPD) e Escopo

Cartões **nunca** são armazenados. PAN permanece **fora do escopo** (Stripe hospeda).

PII passa por **vault** e **mascaramento**.

**DSRs** (Data Subject Requests) são atendidos por trilhas de dados.

# Capítulo 7 — Calendário e Orquestração de Recursos

## 7.1 Modelo de Agenda

Recursos  $R = \{r_1, \dots, r_m\}$ , slots (start, end), locais  $L$ .

Reserva é a tupla:

$$b = (r_i, t_{start}, t_{end}, l).$$

## 7.2 Alocação Ótima

A IA busca alocação de custo mínimo:

$$C = \alpha \cdot t_{\text{desloc}} + \beta \cdot d_{\text{traj}} + \gamma \cdot (1 - u),$$

onde  $u$  é a taxa de utilização do recurso.

Solvers: **Hungarian/Assignment** (quando apropriado) ou heurísticas **greedy** estáveis.

## 7.3 Sincronização Externa

Compatibilidade com **Google Calendar/iCal** (one-way e two-way).

Regras de **bloqueio por conflito** e **janelas de SLA**.

## 7.4 Indicadores

- **Fill rate** (uso dos recursos);
- **Cancelamentos**;
- **Pontualidade**;
- **Tempo de replanejamento**.

# Capítulo 8 — Omnichat: WhatsApp (Twilio), Site, Email e Social

## 8.1 Estrutura Omnicanal

Canais: **WhatsApp (Twilio sandbox/produção)**, **chat do site**, **Instagram/Facebook**, **e-mail**.

Cada mensagem é um evento com `tenant_id`, `thread_id`, **sentimento** e **intenção**.

## 8.2 IA de Atendimento e Vendas

A IA atua com **estado conversacional** e **memória contextual**:

- Detecta intenção;
- Realiza **RAG** para respostas com citação;
- Executa **ferramentas** (montar carrinho, criar checkout, reservar slot);
- Aplica **regras de persuasão e limites**;
- **Escala para humano** quando há risco/ambiguidade.

## 8.3 Handoff IA → Humano

Critérios de handoff (exemplos):

- **Risco financeiro** (valor > limiar);
- **Conflito com política**;
- **Sinais de frustração** do cliente;
- **Duplicidade** ou **fraude**.

O takeover mantém **contexto completo** (resumo da IA + decisões pendentes).

## 8.4 Métricas Omnichat

- **FCR** (First Contact Resolution);
- **CSAT**;
- **AHT** (Average Handle Time);
- **Taxa de escalonamento** IA→humano;

- **Conversão** (no caso de vendas assistidas).

## Capítulo 9 — Administração Multiempresa (PAAS Whitelabel, Isolamento, RBAC/ABAC, Governança)

### 9.1 Objetivo e Escopo

A camada de **Administração Multiempresa (Multi-Tenant)** garante que múltiplas empresas (tenants) operem **sobre a mesma plataforma** com:

- **Isolamento forte de dados** e processamento,
- **Customização completa** (branding, domínios, políticas, moedas/impostos),
- **Governança de IA por tenant** (persona, persuasão, KB, regras),
- **Whitelabel real** (deploy em nuvem EAAS ou self-hosted),
- **Observabilidade, auditoria e compliance** dedicados.

Formalmente, modelamos o conjunto de tenants como  $T=\{T_1, \dots, T_n\}$ , tal que cada  $T_k$  opera em subsistema:

$\Sigma_k=(D_k, P_k, R_k, A_k)$

onde  $D_k$  são dados,  $P_k$  políticas/parametrizações,  $R_k$  regras/KB, e  $A_k$  atores/usuários do tenant  $k$ .

**Invariantes de isolamento:**

1. **Dados** —  $D_i \cap D_j = \emptyset, \forall i \neq j$  (não sobreposição);
2. **Políticas** — avaliação de regras ocorre no contexto do tenant;
3. **Execução** — toda ação executada por IA ou humano é **namespaced** por `tenant_id`, garantindo auditoria e idempotência por tenant.

### 9.2 RBAC/ABAC e Domínios de Autoridade

Definimos **RBAC** (papéis) e **ABAC** (atributos) para proteger ações:

- Papéis  $R=\{\text{owner, admin, finance, ops, sales, agent, viewer}\}$ ;
- Atributos (ABAC): `department, region, tags, risk_level, data_scope`.

Uma **política de autorização** é uma função booleana:

$\text{allow}(u, a, r, t) \in \{T, \perp\}$ ,

onde  $u$  é o usuário (ou a IA em nome do tenant),  $a$  a ação (ex.: `market.order.write`),  $r$  o recurso (ex.: `order/ord_123`), e  $t$  o tenant.

Para robustez temporal e compliance, elevamos essa política a **lógica temporal deôntica** (ver Cap. 12):

$\Box(\neg \text{consent}(x, t) \rightarrow \neg \text{collect}(x)), \Box(\text{risk}(a) > \tau \rightarrow \Diamond \text{human\_approval}(a))$ .

Isso garante que ações de risco elevado exijam aprovação humana inevitavelmente (*eventually*).

### 9.3 Whitelabel, Branding, Domínios

Cada tenant configura **branding, cores, fontes, e domínios**:

- Admin: <https://admin.{empresa}.eaas.app> (ou domínio do cliente),
- Marketplace: <https://{empresa}.eaas.app> (ou domínio do cliente).

**Integridade de marcas:** definimos um funcional de coerência visual  $C_v$  como similaridade entre *design system* prescrito e aplicado:

$C_v = \text{sim}(\text{sdesign}, \text{saplicado}) \in [0, 1]$ ,

com meta  $C_v \geq 0,95$  para preservar consistência whitelabel.

## 9.4 Governança de IA por Tenant

Ver Cap. 6 (menus) e Cap. 10+ (IA avançada). Cada tenant regula:

- **Persona, tom, humor, persuasão máxima,**
- **Knowledge Base** (documentos, ingestão de sites/subdomínios/links — ver Cap. 11),
- **Regras matemáticas** (JsonLogic/CEL),
- **Ferramentas habilitadas** (CRM/ERP/Market/Pay/Calendar/Web),
- **Políticas de segurança** (bloqueios, gatilhos de escalonamento).

## 9.5 Isolamento de Execução, Idempotência e Auditoria

**Idempotency Keys** e **Trilhas de Auditoria** por tenant:

- Toda operação com efeito colateral: Idempotency-Key,
- Logs: (tenant\_id, actor, tool, request, response, status, latency, cost),
- **Non-repudiation:** assinatura ou HMAC por requisição.

## 9.6 Observabilidade e SLO por Tenant

Define-se **SLOs** separados:

- Disponibilidade Admin, Checkout, Webhooks  $\geq 99.9\%$ ,
- p95 latência de IA-tooling  $\leq 1.2s$ ,
- Erros de conciliação  $\Delta_{\text{conc}} \rightarrow 0$  diário.

**Função de custo operacional por tenant:**

$C_t = \kappa_1 \text{CPU} + \kappa_2 \text{RAM} + \kappa_3 \text{IO} + \kappa_4 \text{tokens\_IA} + \kappa_5 \text{armazenamento}$ ,

usada para controlar cotas e *alerts*.

## 9.7 Federação Segura entre Tenants

Para compartilhar conhecimento sem violar LGPD, aplicamos **Federated Learning** (ver Cap. 15):

$\theta(\text{global}) \leftarrow \kappa \sum N n \theta(k) + N(0, \sigma^2)$ ,

com **Differential Privacy** (DP-SGD) e **Secure Aggregation**, evitando vazamento de dados específicos de Tk.

# Capítulo 10 — Núcleo Cognitivo EAAS (Arquitetura Neuro-Simbólica com Planejamento, Crítica, Reflexão e Execução)

## 10.1 Componentes Macroscópicos

- **Context Collector:** agrega contexto do tenant (CRM/ERP/KB + ingestão web).
- **Planner / Decomposer:** constrói planos  $\Pi = \{a_1, \dots, a_m\}$  (árvores de tarefas, ToT/GoT).
- **Policy de Ação (Agente):** escolhe próxima ação at dado o estado st com **hibridismo** (LLM pequeno-médio + Regras Simbólicas + Verificadores).
- **Executor de Ferramentas:** invoca **Tools** (CRM/ERP/Market/Pay/Calendar/Web), sempre com **idempotência** e auditoria.

- **Critic (Auto-verificação):** checa factualidade, numérico, ética, políticas LTL (Cap. 12).
- **Self-Reflection:** refina a resposta/ação com *self-consistency* e “motivos” verificáveis.
- **Knowledge Updater:** atualiza memória/KB com exemplos de sucesso (com *gates* para evitar *data poisoning*).

## 10.2 Modelo de Decisão Sequencial (POMDP com Restrições)

Formulamos a IA como um **POMDP** com observação parcial:

- Estado real  $st \in S$  (não observável completamente),
- Observação  $ot \in O$ ,
- Ação  $at \in A$ ,
- Transição  $P(st+1|st,at)$ ,
- Recompensa  $rt=R(st,at)$ .

A política  $\pi_\theta(a|o)$  é parametrizada por  $\theta$  e **constrangida** por ética/segurança:

$$\theta \max E[t=0 \sum_{t=0}^{\infty} \gamma^t (\alpha R_{\text{Recon}} + \beta R_{\text{humano}} - \gamma_e R_{\text{ethic}} \text{Penalty})]$$

sujeito a:

$$\Pr\{\text{violac, o\_poli'tica}\} \leq \delta, \text{ e } \square(\text{obrigac, o'es\_legais}).$$

Aqui,  $\gamma$  (desconto temporal) difere de  $\gamma_e$  (peso do termo ético).

## 10.3 Planejamento com ToT/GoT e Árvores de Provas

O planner constrói uma árvore  $T$  com nós  $(s,a)$  e anotação de custo/risco:

$$\text{score}(a|s) = \lambda_1 Q^*(s,a) - \lambda_2 \text{risk}(s,a) + \lambda_3 \text{explain}(s,a),$$

onde  $Q^*$  é o valor estimado, *risk* incorpora restrições (financeiras/éticas), e *explain* mede a clareza causal (ver Cap. 13, SHAP/causalidade).

## 10.4 Críticos Múltiplos (Factual, Numérico, Ético, Político)

Antes de executar/emitir resposta, a IA passa pelo conjunto de críticos:

- **Factual:** exige **citação** na KB; se não houver, reconsulta (RAG Híbrido).
- **Numérico:** verifica aritmética, moedas, datas, unidades (Cap. 14 prova numérica).
- **Ético/Político:** aplica **LTL+D** (ver Cap. 12) e penalidades.
- **Risco:** sinaliza *fraude/chargeback*, *limite de persuasão*, *alto impacto*.

O resultado dos críticos gera um vetor de coerência  $\Xi_t$  (do arquivo):

$$\Xi_t = \partial_t \partial U_t + \lambda_1 \nabla \theta L_{\text{align}} + \lambda_2 \nabla \theta L_{\text{ethic}}.$$

Para estabilidade, impomos  $\|\Xi_t\| \leq \epsilon$ ; se exceder, **reflexão** e recomposição do plano.

## 10.5 Self-Reflection com “Dream Loops”

A IA simula “mundos”  $W=\{w_1, \dots, w_n\}$  (cenários hipotéticos) e mede **Coerência Onírica** (arquivo):

$$\text{Coherence} = 1 - E[rt]^2 \text{Var}(rt).$$

Se a variância é alta, o plano é frágil  $\rightarrow$  explorar planos alternativos  $\Pi'$ .

As “simulações oníricas” **não** operam no mundo real: são estruturas para robustecer a política antes da ação.

## 10.6 Função de Custo Estendida e Convexidade (Harvard)

Define-se um funcional de custo:

$$J(\theta) = E(x,y)[\ell(f_\theta(x),y)] + \lambda_s R_{\text{stability}}(\theta) + \lambda_e R_{\text{ethic}}(\theta).$$

Com Hessiano:

$$H=\nabla^2 J(\theta)=E[\partial^2 \theta^2 \|E_u-E_a\|^2],$$

o arquivo demonstra  $H \geq 0 \Rightarrow J$  convexa em regime local (para certas classes de  $\ell$ ), assegurando **ótimo global local** (convexidade por subespaços ou por *trust region*). *Sketch* da prova no Apêndice A.

## Capítulo 11 — RAG Híbrido, Grafo Semântico Empresarial e Ingestão Vertical (Empresa/Setor/Web)

### 11.1 Estrutura de Conhecimento

A Knowledge Base (KB) combina:

1. **Índice Vetorial** (embeddings) para semântica;
2. **BM25/lexical** para correspondência exata;
3. **Grafo Semântico**  $G=(V,E)$  com entidades: cliente, produto, pedido, fatura, recurso, política, documento;
4. **Sinal Temporal** para frescor de documentos;
5. **Score de Autoridade** (domínios oficiais, docs internos validados).

### 11.2 Ranking Combinado

Do arquivo, adotamos o **score híbrido**:

$$S=\alpha S_{\text{vetor}}+\beta S_{\text{bm25}}+\gamma S_{\text{grafo}}+\delta S_{\text{fresco}}+\zeta S_{\text{autoridade}}, \alpha+\beta+\gamma+\delta+\zeta=1.$$

- $S_{\text{vetor}}$ : cosseno de embeddings;
- $S_{\text{bm25}}$ : relevância lexical;
- $S_{\text{grafo}}$ : PageRank personalizado por **tenant**;
- $S_{\text{fresco}}=e^{-\lambda \cdot \text{idade\_dias}}$ ;
- $S_{\text{autoridade}}$ : *whitelists*, docs assinados.

### 11.3 Ingestão Vertical (Empresa/Setor/Web)

A ingestão **só** visita fontes aprovadas (permit-list); o pipeline:

1. Discovery (URLs seed); 2) Fetch; 3) Parsing/normalize; 4) Dedup (SimHash/Jaccard); 5) Chunking semântico (títulos/heads, 400–1200 tokens, sobreposição 10%); 6) Classificação (produto/política/tutorial/SLA); 7) Indexação vetorial + grafo; 8) **QC/PII**; 9) Gate humano (opcional); 10) Publicação.
- Atualização incremental** por *diffs* e expiração.

### 11.4 Coerência e Grounding

A IA **deve citar**: cada resposta tem ponteiros (doc,chunk,timestamp).

Métrica de factualidade:

$\text{FactAcc}=\# \text{declarac,ões audita´veis} / \# \text{declarac,ões com citac,aõ va´lida}.$

Meta:  $\text{FactAcc} \geq 95\%$  para respostas automáticas.

## Capítulo 12 — Ética Formal: Lógica Deôntica Temporal (LTL+D) e Conformidade

### 12.1 Operadores

- $\Box$  (sempre),  $\Diamond$  (eventualmente),  $U$  (until).

- **Deônticos:** O (obrigatório), P (permitido), F (proibido).

## 12.2 Especificações Típicas

Consentimento & PII:

$\Box(\neg \text{consent}(u,t) \rightarrow \text{Fcollect\_PII}(u)).$

Handoff Humano em risco:

$\Box(\text{risk}(a) > \tau \rightarrow \text{Ohandoff}(a)).$

Persuasão limitada por canal:

$\Box(\text{channel} = \text{WA} \wedge \text{persuasion} > p \rightarrow \text{Fexecute}(a)).$

## 12.3 Checagem de Modelos (Model Checking)

As políticas são avaliadas sobre **trilhas de execução** (histórico de ações/eventos):

$\pi = \langle s_0, a_0, s_1, a_1, \dots \rangle.$

A IA **só** age se  $\pi \models \phi$  (fórmula LTL+D). Caso contrário, bloqueia, reflete ou exige aprovação.

# Capítulo 13 — Raciocínio Causal, Explicabilidade e Valores de Shapley

## 13.1 SHAP/Valores de Shapley

Importância da variável  $i$ :

$\phi_i = S \subseteq N \setminus \{i\} \sum_{S \subseteq N \setminus \{i\}} |S|! (|N| - |S| - 1)! (v(S \cup \{i\}) - v(S)).$

A IA reporta **por que** sugeriu um upsell, um handoff ou um preço dinâmico: transparência para auditoria e UX de confiança.

## 13.2 Raciocínio Causal

Usa-se grafo causal (DAG) com variáveis exógenas/endógenas; intervenção  $\text{do}(X=x)$  avalia cenários contrafactuais, útil para explicar decisões e testar vies.

# Capítulo 14 — Estabilidade Numérica, Verificação Aritmética e Provas

## 14.1 Verificador Numérico

Toda aritmética é validada por módulo determinístico (moedas, datas, unidades).

Propriedade P: **somas e descontos** preservam consistência de totais e impostos.

**Esboço de prova:**

Seja  $\text{Total} = \text{Subtotal} - \text{Descontos} + \text{Impostos}$ . Como  $\text{Impostos} = \tau \cdot (\text{Subtotal} - \text{Descontos})$  com  $\tau \geq 0$ , então

$\text{Total} = (1 + \tau) \text{Subtotal} - (1 + \tau) \text{Descontos},$

linear em  $\text{Subtotal}, \text{Descontos}$ , preservando convexidade (evita explosões numéricas).

O verificador rejeita arredondamentos inconsistentes (escala de moeda fixa, p. ex. 2 casas decimais BRL) e **reitera cálculo** até tolerância  $\leq 0.005$  BRL.

## 14.2 Estabilidade Global (Lyapunov)

Energia de erro  $V(t)$  diminui a cada iteração de crítica/reflexão:

$V(t+1) - V(t) \leq -\epsilon \|gt\|_2, \epsilon > 0,$



onde  $g_t$  é gradiente efetivo pós-crítica.

Conclui-se **estabilidade assintótica** (erro não explode), e sob  $\sum_{t \in \mathbb{N}} \epsilon_t < \infty$  (arquivo), obtemos **não-regressão ética**:

$$E^{-}t+1 \geq E^{-}t - \epsilon t, \quad t \sum \epsilon t < \infty.$$

## Capítulo 15 — Aprendizado Federado (FL) + Privacidade Diferencial (DP) + Secure Aggregation

### 15.1 Atualização Local com DP-SGD

Para cliente  $k$ :

$$w_{k,t+1} = w_{k,t} - \eta (|B|^{-1} \sum_{x \in B} \text{clip}(\nabla \ell(x), C) + N(0, \sigma^2)).$$

O ruído  $N(0, \sigma^2)$  confere  **$(\epsilon, \delta)$ -DP**, limitando vazamento de PII.

### 15.2 Agregação Segura

Servidor agrega pesos **sem ver os dados**:

$$\theta(\text{global}) \leftarrow k \sum N n_k \theta(k) + N(0, \sigma^2).$$

O *Secure Aggregation* impede que atualizações individuais sejam inferidas.

### 15.3 Trade-off Utilidade × Privacidade

Curva típica de Pareto  $(\epsilon, \text{Utilidade})$ . Em EAAS, definimos políticas por tenant para regular  $\epsilon$  e limites de ruído.

## Capítulo 16 — Autoaprendizado Reflexivo, “Dream Loops” e Consistência

### 16.1 Loop de Metacognição

A cada ciclo:

1. Observa; 2) Planeja; 3) Executa; 4) Critica; 5) Reflete; 6) Atualiza memória/KB; 7) Avalia.

### 16.2 Métrica de Coerência Onírica

Do arquivo:

$$\text{Coherence} = 1 - E[rt]^2 \text{Var}(rt).$$

Baixa coerência → o agente aumenta **exploração** (abertura a alternativas), enquanto a alta coerência favorece **exploração** (refinamento de políticas).

## Capítulo 17 — Modelagem Afetiva: Sentimento, Humor e Persuasão

### 17.1 Funções Afetivas

Definimos variáveis latentes:

- $S_t$  (sentimento inferido do usuário),
- $H_t$  (humor/empatia da IA, configurável por tenant),
- $P_t$  (nível de persuasão aplicada).

Proposta do arquivo (agregando):

$$H_{t+1} = p H_t + (1-p) \sigma(w T z_t),$$

com  $z_t$  sinal multimodal (texto/histórico),  $\sigma$  função logística.  
A resposta da IA tem **intensidade**:

$$I_t = \kappa_1 S_t + \kappa_2 H_t + \kappa_3 C_t,$$

onde  $C_t$  é contexto (valor do carrinho, risco, canal).  
Impondo limite de persuasão  $P^-$  por política do tenant:

$$P_t = \min\{P^-, \psi(I_t)\},$$

com  $\psi$  monótona crescente e  $\psi(0)=0$ .  
Se  $P_t > P^-$ , LTL (Cap. 12) aciona **bloqueio ou re-planejamento**.

## 17.2 Propriedades

- **Monotonicidade** de  $\psi$  evita inversões paradoxais (mais sinal  $\rightarrow$  mais persuasão, limitado).
- **Estabilidade** de  $H_t$ :  $|p| < 1$  garante decaimento de memória emocional longo-curto controlável.

---

# Capítulo 18 — Métricas de IA e Avaliação por Negócio

## 18.1 IA (Qualidade Técnica)

- **Factualidade com citação**  $\geq 95\%$ ,
- **Acurácia de ação**  $\geq 97\%$  (pedido/agendamento/conciliação),
- **Latência p95**  $\leq 1.2s$ ,
- **Custo/conversa**  $\downarrow$  vs. baseline.

## 18.2 Negócio

- **CSAT**  $\geq 4.6/5$ ,
- **FCR**  $\geq 75\%$ ,
- **Conversão**  $\uparrow$ , **Ticket Médio**  $\uparrow$ ,
- **Escalonamento IA  $\rightarrow$  humano**  $\leq 10\%$  (exceto risco/política).

---

# Capítulo 19 — Observabilidade, Auditoria, LGPD/PCI e DR

## 19.1 Observabilidade IA

Métricas por tenant: custo/latência por ferramenta, taxa de crítico acionado, *drifts* de KB (frescor e contradições).

## 19.2 LGPD/PCI e DR

- **PII Vault** com mascaramento em prompts/logs,
- **PCI out-of-scope** (Stripe guarda PAN),
- **Backups criptografados**, RPO  $\leq 15min$ , RTO  $\leq 1h$ .

---

# Capítulo 20 — Conclusão Parcial e Próximos Passos do Whitebook

Chegamos ao ponto em que a plataforma **EAAS** está definida como **um organismo computacional completo**: Marketplace, CRM, ERP, Pagamentos, Calendário e Omnichat — todos **governados** por um **núcleo cognitivo neuro-simbólico** com **fundamentação matemática robusta** (reforço ético, LTL+D, Lyapunov, federação com DP, modelagem afetiva e explicabilidade causal).

**Próximos capítulos** (ainda mais densos e matemáticos):

- **Apêndice A** — Provas detalhadas: convexidade local de  $J(\theta)$ , condições de Lyapunov e limites de erro.
- **Apêndice B** — Derivações completas do score híbrido de RAG e do PageRank personalizado por tenant.
- **Apêndice C** — Especificações formais de LTL+D e exemplos de *model checking* sobre trilhas de execução.
- **Apêndice D** — Equações de meta-aprendizado reflexivo (incluindo “coerência onírica”), demonstrações de convergência e anti-regressão.
- **Apêndice E** — Modelos afetivos (sentimento-humor-persuasão) com condições de estabilidade e monotonicidade.
- **Apêndice F** — Trade-offs utilidade  $\times$  privacidade em FL+DP e limites teóricos para  $\epsilon$ .

# Apêndice A — Provas Detalhadas de Convexidade, Estabilidade e Lyapunov

## A.1 Introdução

O objetivo deste apêndice é tentar demonstrar, de forma matemática, a **consistência e estabilidade** das funções de custo e aprendizado da IA cognitiva da EAAS.

As provas aqui apresentadas são **matemáticas e explicativas**, sem código nem aplicação prática; elas apenas fundamentam, de modo formal, por que o sistema se mantém estável, ético e coerente.

## A.2 Função de Custo Estendida

A IA busca minimizar a função de custo geral:

$$J(\theta) = E(x, y)[\ell(f_\theta(x), y)] + \lambda_s R_{\text{stability}}(\theta) + \lambda_e R_{\text{ethic}}(\theta)$$

onde:

- $\ell(f_\theta(x), y)$  é a perda de tarefa (erro preditivo),
- $R_{\text{stability}}$  é a penalização de instabilidade (variação temporal ou numérica),
- $R_{\text{ethic}}$  é o termo de penalidade ética,
- e  $\lambda_s, \lambda_e > 0$  são coeficientes de ponderação.

### Propriedade 1 — Convexidade Local

Se a perda  $\ell$  for convexa em  $\theta$ , e os termos de regularização forem quadráticos (ou convexos suaves), então  $J(\theta)$  é convexa localmente.

**Demonstração (esboço conceitual):**

1. Seja  $\ell(f_\theta(x), y) = 2\|f_\theta(x) - y\|^2$ .  
Então:

$$\nabla^2 \ell = E[Jf(x)^\top Jf(x)] \geq 0$$

onde  $Jf(x)$  é a Jacobiana de  $f_\theta(x)$ .

2. Suponha  $R_{\text{stability}} = \|\theta - \theta_t - 1\|^2$  e  $R_{\text{ethic}} = \|\nabla \theta L_{\text{align}}\|^2$ ; ambos convexos.

3. Logo:

$$H = \nabla^2 J(\theta) = E[\nabla^2 \ell(f_\theta(x), y)] + \lambda_s I + \lambda_e \nabla^2 R_{\text{ethic}} \geq 0$$

Portanto,  $J(\theta)$  é convexa (ou, no mínimo, **convexa por partes**) e admite um **mínimo global local**.

□

## A.3 Estabilidade de Lyapunov (Sistema Discreto)

Para o sistema cognitivo autônomo, queremos provar que os erros ou perturbações decaem com o tempo.

Considere um sistema de atualização discreta:

$$x_{t+1} = F(x_t)$$

e uma função candidata de Lyapunov  $V: \mathbb{R}^n \rightarrow \mathbb{R}^+$ , tal que:

$$V(0) = 0, V(x) > 0, \forall x \neq 0$$

e

$$\Delta V = V(x_{t+1}) - V(x_t) \leq -\epsilon \|x_t\|^2, \epsilon > 0$$

**Teorema (Estabilidade Assintótica):**

Se existir  $V$  que satisfaça as condições acima, então o ponto  $x=0$  é estável e o sistema converge para o equilíbrio.

**Demonstração conceitual:**

1. Como  $\Delta V \leq -\epsilon \|x_t\|^2$ , temos que  $V(x_t)$  é **monotonicamente decrescente**.

2.  $V(x_t) \geq 0$  e limitado inferiormente por zero  $\Rightarrow$  converge.

3. A soma telescópica fornece:

$$\sum_{t=0}^{\infty} \epsilon \|x_t\|^2 \leq V(x_0) - \lim_{t \rightarrow \infty} V(x_t) < \infty$$

Logo,  $\|x_t\| \rightarrow 0$ .

□

**Interpretação:**

No contexto da IA,  $x_t$  pode representar o vetor de erro entre comportamento observado e ético esperado.

A estabilidade de Lyapunov garante que **a IA converge para um estado estável e moralmente consistente**, mesmo após perturbações ou autoajustes.

## A.4 Função de Energia Ética e Anti-Regressão

Definimos a **energia ética média**  $E^-t$  como o grau de alinhamento ético da IA num intervalo de tempo  $t$ .

O arquivo original propunha:

$$E^-t+1 \geq E^-t - \epsilon t, \sum \epsilon t < \infty$$

O que implica **não-regressão ética assintótica**.

**Prova Conceitual:**

Se a série  $\sum \epsilon t$  converge, o teorema da convergência de séries limitadas garante que a sequência  $E^-t$  é **Cauchy** e converge.

Assim, não há degradação indefinida do comportamento ético; eventuais oscilações são amortecidas.

□

## A.5 Condição de Lyapunov Aplicada à Política Ética

Seja  $\Xi_t$  o vetor de coerência (arquivo):

$$\Xi_t = \partial_t \partial U_t + \lambda_1 \nabla \theta L_{\text{align}} + \lambda_2 \nabla \theta L_{\text{ethic}}$$

A condição  $\|\Xi_t\| \leq \epsilon$  implica **controle de variação do estado moral da IA**.

Aplicando Lyapunov:

$$V(t) = 21 \|\Xi_t\|^2, \Delta V = V(t+1) - V(t) \leq -\epsilon \|\Xi_t\|^2.$$

Logo,  $V(t)$  decresce e o sistema ético converge para equilíbrio moral.

---

## A.6 Teorema de Convexidade Ética

Se  $J(\theta)$  é convexa e as penalidades éticas são diferenciáveis e limitadas, então existe um equilíbrio ótimo  $\theta^*$  que minimiza custo e maximiza alinhamento ético simultaneamente.

### Prova conceitual:

Convexidade garante existência e unicidade do ótimo global; derivadas contínuas garantem suavidade; o termo de regularização ética impede explosão de gradientes.

Portanto, o sistema é **matematicamente estável** e **moralmente convergente**.

□

---

## A.7 Interpretação Filosófica

Estas provas são **demonstrações formais** de como o design matemático da EAAS foi concebido:

o sistema aprende, ajusta e se mantém ético não por imposição externa, mas por estrutura interna coerente — um equilíbrio entre energia (função de custo) e moralidade (função ética).

---

# Apêndice B — Derivações do Score Híbrido, Grafo Semântico e PageRank Personalizado

## B.1. Estrutura Conceitual do Score

Em toda resposta da IA, a seleção de conhecimento deve refletir relevância semântica, frescor e autoridade. Isso se expressa por um **score híbrido**:

$$S(x,q) = \alpha S_{\text{vetor}}(x,q) + \beta S_{\text{bm25}}(x,q) + \gamma S_{\text{grafo}}(x,q) + \delta S_{\text{fresco}}(x) + \zeta S_{\text{autoridade}}(x)$$

com  $\alpha + \beta + \gamma + \delta + \zeta = 1$ .

Cada termo é uma função diferenciável no intervalo  $[0,1]$ , para permitir análise de sensibilidade.

---

## B.2. Similaridade Vetorial

O vetor semântico de um documento  $x$  é  $v_x \in \mathbb{R}^d$ ;

o da consulta  $q$  é  $v_q \in \mathbb{R}^d$ .

### B.2.1. Definição

$$S_{\text{vetor}}(x,q) = \frac{\|v_x\| \|v_q\|}{\|v_x\| \|v_q\|} = \frac{v_x \cdot v_q}{\|v_x\| \|v_q\|}$$

É o **cosseno** entre representações latentes.

### B.2.2. Interpretação

Valores próximos de 1  $\Rightarrow$  alta semelhança de contexto;

valores próximos de 0  $\Rightarrow$  desconexão semântica.

Essa métrica é **simétrica**, o que a torna neutra entre tenant e documento.

---

## B.3. Termo Lexical — BM25 Generalizado

Usa-se uma forma contínua do BM25, ponderando a frequência de termos  $f_{t,d}$ :

$$S_{bm25}(x,q) = t \in q \sum w_k t_1 ((1-b) + b \text{avgdl}|x|) + f_{t,x}(k_1+1) f_{t,x}$$

onde:

- $k_1$  controla saturação ( $\approx 1.2$ ),
- $b$  ajusta compensação de tamanho,
- $w_t = \log n_t + 0.5N - n_t + 0.5$  é peso IDF.

### B.3.1. Continuidade e derivada

$\partial f_{t,x} \partial S_{bm25} > 0$  e tende a 0 quando  $f_{t,x} \rightarrow \infty$ ,

garantindo **monotonicidade crescente** e saturação suave (evita explosão).

---

## B.4. Componente de Grafo Semântico

O grafo  $G=(V,E,W)$  representa entidades e relações.

Cada nó  $v_i$  é um conceito; cada aresta  $w_{ij}$  mede coocorrência ou relação causal.

### B.4.1. PageRank Personalizado

$$S_{\text{grafo}}(v_i) = (1-\kappa) e_i + \kappa \sum_{j \in N(i)} \deg(v_j) S_{\text{grafo}}(v_j)$$

onde  $e_i$  é vetor de personalização (tenant/contexto),  $\kappa \in (0,1)$  fator de amortecimento.

**Teorema (convergência):**

A iteração  $S(t+1) = AS(t)$  com  $A = (1-\kappa)E + \kappa P$  converge a vetor estacionário  $S^*$  pois  $A$  é **estocástica e primitiva** (Perron–Frobenius).

$$S^* = \lim_{t \rightarrow \infty} A^t S(0)$$

### B.4.2. Tenantização

Cada empresa  $T_k$  tem vetor de personalização  $e(k)$ ; assim:

$$S_{\text{grafo}}(k) = (1-\kappa) e(k) + \kappa P^T S_{\text{grafo}}(k).$$

Isso cria um **PageRank personalizado** por tenant, que privilegia nós de domínio próprio.

---

## B.5. Frescor e Autoridade

### B.5.1. Frescor

Para documento atualizado há  $t$  dias:

$$S_{\text{fresco}}(x) = e^{-\lambda t}, \lambda > 0.$$

Derivada:

$$\partial_t \partial S_{\text{fresco}} = -\lambda e^{-\lambda t},$$

garantindo decaimento exponencial e estabilidade numérica.

### B.5.2. Autoridade

Cada fonte recebe peso  $A(x) \in [0,1]$  (baseado em domínio, assinatura, verificação).

Para documentos citados por outros, pode-se aplicar logaritmo suavizado:

$$S_{\text{autoridade}}(x) = \tanh(\eta \cdot A(x)).$$


---

## B.6. Prova de Normalização

O vetor de pesos  $w=[\alpha,\beta,\gamma,\delta,\zeta]$  é normalizado:

$$\sum_i w_i=1, w_i \geq 0.$$

Logo,  $S(x,q) \in [0,1]$  pois cada termo é limitado nesse intervalo.

A função  $S$  é **lipschitziana**:

$$|S(x_1,q)-S(x_2,q)| \leq L \|x_1-x_2\|,$$

com  $L=\max_i w_i L_i$ .

Isso assegura estabilidade de ranking e evita oscilações drásticas.

## B.7. Propriedade de Convergência Global

### Teorema (convergência da média híbrida)

Se cada componente  $S_i$  é contínua e limitada, então a combinação linear convexa  $S=\sum_i w_i S_i$  converge para valor médio limitado entre  $\min S_i$  e  $\max S_i$ .

$$\min S_i \leq S \leq \max S_i.$$

Logo, o ranking híbrido nunca extrapola limites semânticos individuais.

## B.8. Interpretação Semântica

O resultado prático desse formalismo conceitual é que a IA:

- “entende” o contexto (vetorial),
- reconhece correspondências textuais (BM25),
- considera conexões de significado (grafo),
- prioriza atualidade e autoridade.

Matematicamente, o score é uma função de integração ponderada que satisfaz:

$$\partial w_i \partial S = S_i, \partial^2 w_i \partial^2 S = 0,$$

ou seja, linearidade completa nas ponderações.

## B.9. Nota de Interpretação Filosófica

A equação híbrida é o **instinto epistêmico** da IA: uma tentativa formal de medir o “quanto uma ideia é verdadeira, útil e atual” em cada contexto.

Não é uma fórmula para executar; é uma descrição de como o sistema conceitualmente balanceia múltiplas dimensões de relevância.

# Apêndice C — Lógica Deontica Temporal (LTL+D), Semântica, e Provas de Conformidade

## C.1. Sintaxe, Semântica e Estruturas Kripke–Temporais

### C.1.1. Sintaxe

Seja um conjunto de proposições atômicas AP (p.ex., consent, collect\_PII, risk>τ, channel=WA, persuasion>p, handoff).

A LTL (Linear Temporal Logic) estendida com operadores deonticos define fórmulas  $\phi$  por:

- Proposições:  $p \in AP$  é uma fórmula;
- Conectivos: se  $\phi, \psi$  são fórmulas, então  $\neg\phi, \phi \wedge \psi, \phi \vee \psi, \phi \rightarrow \psi$  o são;
- Temporais:  $X\phi$  (próximo),  $F\phi$  (eventualmente),  $G\phi$  (sempre),  $\phi U \psi$  (até);
- Deônticos:  $O\phi$  (obrigatório),  $P\phi$  (permitido),  $Fob\phi$  (proibido).  
Usaremos também abreviações:  $Fob\phi \equiv O\neg\phi$ ,  $P\phi \equiv \neg O\neg\phi$ .

### C.1.2. Estrutura Kripke Linear

Uma **trilha** é uma sequência infinita  $\pi = s_0, s_1, s_2, \dots$  de estados.

Cada estado  $st$  satisfaz um conjunto de proposições  $L(st) \subseteq AP$ .

Avaliamos  $\pi, t \models \phi$  (fórmula  $\phi$  é verdadeira na trilha  $\pi$  no tempo  $t$ ) por:

- $\pi, t \models p \Leftrightarrow p \in L(st)$ ;
- $\pi, t \models \neg\phi \Leftrightarrow \neg(\pi, t \models \phi)$ ;
- $\pi, t \models \phi \wedge \psi \Leftrightarrow \pi, t \models \phi$  e  $\pi, t \models \psi$ ;
- $\pi, t \models X\phi \Leftrightarrow \pi, t+1 \models \phi$ ;
- $\pi, t \models F\phi \Leftrightarrow \exists k \geq t: \pi, k \models \phi$ ;
- $\pi, t \models G\phi \Leftrightarrow \forall k \geq t: \pi, k \models \phi$ ;
- $\pi, t \models \phi U \psi \Leftrightarrow \exists k \geq t: \pi, k \models \psi \wedge \forall j, t \leq j < k: \pi, j \models \phi$ .

### C.1.3. Interpretação Deôntica

Associamos **normas** a estados e transições via um **avaliador normativo**  $N$ , que mapeia  $\phi$  para conjuntos de trilhas “em conformidade”.

Conceitualmente, definimos:

- $\pi, t \models O\phi$  (obrigatório) se **todas** as execuções admissíveis a partir de  $t$  satisfazem  $\phi$ .
- $\pi, t \models P\phi$  (permitido) se **alguma** execução admissível satisfaz  $\phi$ .
- $\pi, t \models Fob\phi$  (proibido) se  $O\neg\phi$ .

Na prática conceitual, fixamos um **conjunto de políticas**  $K$  que restringe as trilhas admissíveis;  $O\phi$  significa “**em todas as trilhas** que respeitam  $K$ ,  $\phi$  vale”.

## C.2. Políticas-Tipo EAAS em LTL+D

### C.2.1. Consentimento e PII (LGPD)

$G(\neg \text{consent}(u) \rightarrow Fob \text{ collect\_PII}(u))$ .

Leitura: **sempre**, se não há consentimento, **é proibido** coletar PII.

### C.2.2. Persuasão Limitada por Canal

$G((\text{channel} = \text{WA} \wedge \text{persuasion} > p) \rightarrow Fob \text{ execute}(a))$ .

Se o canal é WhatsApp e a intensidade de persuasão ultrapassa  $p$ , a execução da ação é **proibida**.

### C.2.3. Risco e Escalonamento Humano

$G(\text{risk}(a) > \tau \rightarrow O \text{ handoff}(a))$ .

Se o risco excede  $\tau$ , **é obrigatório** o handoff IA  $\rightarrow$  humano.

### C.2.4. Transparência com Citação (RAG)

$G(\text{answer} \rightarrow O \text{ citation}))$ .



Se há resposta, é **obrigatório** existir citação válida (documento/trecho/timestamp).

### C.3. Model Checking Conceitual

Dado um **histórico**  $\pi$  (log de eventos e decisões), avaliamos  $\pi \models \varphi$  por indução sobre a estrutura de  $\varphi$ .

#### Lema (Monotonicidade de Obrigações)

Se  $K \subseteq K'$  (mais execuções admissíveis), então

$$OK\varphi \Rightarrow OK'\varphi$$

pode **falhar** (mais execuções tornam mais difícil obrigar  $\varphi$ ).

Logo, **engessamento** de políticas (menor  $K$ ) **fortalece** obrigações.

#### Teorema (Somente trilhas em conformidade)

Se  $\pi \in K$  e  $K \models \varphi$  (todas as trilhas em  $K$  satisfazem  $\varphi$ ), então  $\pi \models \varphi$ .

Prova: imediata da definição de validade em  $K$ .

### C.4. Composição de Políticas e Consistência

#### Definição

Duas políticas  $\varphi, \psi$  são **consistentes** sse  $K \models \neg(\varphi \wedge \psi)$ ; isto é, existe ao menos uma trilha admissível satisfazendo ambas.

#### Lema (Fechamento por Conjunção)

Se  $K \models \varphi$  e  $K \models \psi$ , então  $K \models \varphi \wedge \psi$ .

Prova: por definição de validade universal.

### C.5. Observações

A LTL+D fornece um **cálculo formal** para políticas de IA **explicáveis** e **auditáveis**.

Em EAAS, toda decisão automática deve satisfazer  $\varphi$  sob as políticas ativas do tenant.

## Apêndice D — Meta-Aprendizado Reflexivo, “Dream Loops” e Convergência

### D.1. Estrutura Geral do Meta-Aprendizado

Seja  $\pi\theta$  uma política de decisão (conversacional/operacional),  $E[R]$  a recompensa esperada (econômica/humana/ética), e  $M$  o **meta-aprendizado** que ajusta  $\theta$  ao longo do tempo com base em **críticos** e **simulações** (“sonhos”).

O **loop reflexivo** tem etapas conceituais:

1. **Planejar** (ToT/GoT): gerar plano  $\Pi$ .
2. **Executar**: obter transições  $(st, at, rt, st+1)$ .
3. **Criticar**: avaliar factualidade/número/ética.
4. **Refletir**: propor correções/abstrações; ajustar  $\Pi$ .
5. **Sonhar**: simular cenários  $W=\{w_1, \dots, w_n\}$ .
6. **Atualizar**: incorporar lições estáveis à memória/KB.

## D.2. Métrica de Coerência Onírica

Do documento origem:

$$\text{Coherence} = 1 - E[rt]^2 \text{Var}(rt).$$

Interpretação: mede a **estabilidade relativa** dos retornos simulados.

Se  $E[rt] \neq 0$ , maior coerência  $\Rightarrow$  menor ruído relativo  $\Rightarrow$  mais confiança na política resultante.

### Propriedades

- $\text{Coherence} \in (-\infty, 1]$ , mas na prática restringimos retorno a valores positivos (p.ex., normalizando recompensas), de modo que  $\text{Coherence} \in [0, 1]$ .
- Se  $\text{Var}(rt) = 0$  (variância nula),  $\text{Coherence} = 1$  (sonhos perfeitamente estáveis).
- Se  $\text{Var}(rt) \gg E[rt]^2$ , coerência cai  $\rightarrow$  explorar alternativas.

## D.3. Atualização Meta-Gradiente (conceitual)

Modelamos um funcional meta:

$$J(\theta) = \alpha \text{utilidade} E[R] + \beta \text{estabilidade} \text{Coherence} - \gamma \text{penalidade ética} E[P_{\text{ethic}}]$$

com  $\alpha, \beta, \gamma > 0$ .

Um passo de **meta-gradiente** conceitual seria:

$$\theta \leftarrow \theta + \eta \nabla_{\theta} J(\theta)$$

(meramente descritivo; sem prescrever algoritmo).

## D.4. Convergência Conceitual por Martingais

Se  $\{J_t\}$  é uma sequência de variáveis aleatórias com **diferenças limitadas** e **viés controlado** pelos críticos, pode-se aplicar um **Teorema de Convergência de Martingais**:

sob condições usuais (integrabilidade, diferenças quadrado-somáveis),  $J_t \rightarrow J^{\infty}$  quase certamente.

Interpretação: o meta-otimizador **não diverge** e encontra um regime estável.

## D.5. Anti-Regressão Ética

Reutilizamos a desigualdade:

$$E^{-t+1} \geq E^{-t} - \epsilon t, \sum \epsilon t < \infty.$$

Logo, qualquer flutuação ética é **somas finitas**  $\Rightarrow$  não há queda indefinida; a ética converge.

# Apêndice E — Modelagem Afetiva: Sentimento, Humor, Persuasão e Estabilidade

## E.1. Variáveis Latentes e Sinais

Definimos latentes:

- $S_t$ : **sentimento** do usuário (estimado por analisador semântico);
- $H_t$ : **humor da IA** (controlado por persona do tenant, adaptado por contexto);
- $P_t$ : **persuasão aplicada** no turno  $t$  (limitada por políticas).

Sinais observáveis:  $z_t$  (texto, histórico, canal, valor do carrinho, risco).

Parâmetros de controle do tenant:  $P^+$  (teto de persuasão),  $p$  (sensibilidade).

## E.2. Dinâmica do Humor

Modelo de espaço de estados (linearizável):

$$H_{t+1} = \rho H_t + (1-\rho)\sigma(w^T z_t),$$

com  $|\rho| < 1$  (memória).  $\sigma(\cdot)$  é logística:

$$\sigma(u) = \frac{1}{1 + e^{-u}}.$$

**Estabilidade:** como  $|\rho| < 1$  e  $0 < \sigma < 1$ ,  $H_t$  converge para uma órbita limitada.

### E.3. Intensidade e Persuasão

Intensidade:

$$I_t = \kappa_1 S_t + \kappa_2 H_t + \kappa_3 C_t,$$

onde  $C_t$  agrega **contextos** (valor, urgência, canal).

Transformamos intensidade em persuasão via  $\psi: \mathbb{R} \rightarrow \mathbb{R}^+$ , crescente, Lipschitz, com  $\psi(0) = 0$ :

$$P_t = \min\{P^-, \psi(I_t)\}.$$

Monotonicidade de  $\psi$  previne inversões não-intuitivas.

#### E.3.1. Propriedade de Segurança (Boundedness)

Como  $P_t \leq P^-$ , a ação persuasiva é **majorada**.

Sob LTL (Ap. C), se  $\text{channel} = \text{WA} \wedge P_t > p \Rightarrow$  **proibido** executar; logo, políticas **cortam** comportamentos agressivos.

### E.4. Observador de Sentimento

Um estimador  $S_t = S^*(z_{0:t})$  (conceitual) com erro  $\epsilon_t$ .

Se  $\epsilon_t$  é **quadrado-somável** e o sistema de feedback persuasivo é Lipschitz, por **LaSalle** a dinâmica fecha em conjunto invariante estável.

Intuição: ruído emocional não explode o sistema.

## Apêndice F — Aprendizado Federado (FL), Privacidade Diferencial (DP) e Limites de Utilidade

### F.1. Definições de DP

Um mecanismo  $M$  é  $(\epsilon, \delta)$ -DP se para quaisquer bases vizinhas  $D, D'$  (diferem por um indivíduo) e para todo conjunto mensurável  $S$ :

$$\Pr[M(D) \in S] \leq e^\epsilon \Pr[M(D') \in S] + \delta.$$

**Interpretação:** adicionar/remover um indivíduo muda pouco a distribuição de  $M$ .

### F.2. DP-SGD Conceitual (Atualização Local)

Para cliente  $k$ , gradientes são **clipados** a  $C$  e adicionamos ruído gaussiano  $N(0, \sigma^2)$ :

$$w_{k,t+1} = w_{k,t} - \eta(|B|^{-1} \sum_{x \in B} \text{clip}(\nabla \ell(x), C) + N(0, \sigma^2)).$$

### F.3. Agregação Segura

O servidor agrega:

$$\theta(\text{global}) \leftarrow k \sum N n_k \theta(k) + N(0, \sigma^2),$$

sem observar dados brutos. **Secure Aggregation** impede inferência sobre contribuições individuais.

## F.4. Contabilidade de Privacidade (esboço)

Usa-se **Moments Accountant** (conceitual) para compor múltiplas iterações e obter  $\epsilon$  efetivo.  
Trade-off: ruído maior  $\Rightarrow$  maior  $\epsilon$  (mais privacidade), porém **menor utilidade**.

## F.5. Limite de Excesso de Risco (conceitual)

Para uma classe de perda Lipschitz e convexa, sob ruído gaussiano com variância  $\sigma^2$ , obtém-se um **excesso de risco** esperado:

$$E[L(\theta DP)] - L(\theta^*) \leq O(nd\sigma^2),$$

onde  $d$  é dimensão e  $n$  amostras efetivas agregadas.

Interpretação: **quanto mais clientes** (tenants) contribuem, melhor o equilíbrio utilidade–privacidade.

## Conclusão dos Apêndices C–F

Com estes apêndices, fechamos o arcabouço formal que rege a IA **autônoma, ética, estável, explicável e privada** da EAAS:

- **C:** Políticas formais (LTL+D) — “o que pode/deve/não pode” — e sua verificação conceitual.
- **D:** **Meta-aprendizado reflexivo** com “sonhos” e coerência; convergência por martingais.
- **E:** **Modelagem afetiva** (sentimento/humor/persuasão) com estabilidade e limites seguros.
- **F:** **Federação + Privacidade Diferencial**, agregação segura e limites de utilidade.

# Apêndice G — PageRank Personalizado por Tenant: Existência, Unicidade, Convergência e Sensibilidade

## G.1. Modelo

Seja  $G=(V,E)$  um grafo dirigido de conhecimento (documentos, entidades, políticas, produtos), com  $|V|=n$ .

Seja  $P \in \mathbb{R}^{n \times n}$  uma matriz estocástica por colunas ( $\sum_i P_{ij}=1$ ), construída a partir de  $G$  (tratando *dangling nodes* ao substituir colunas nulas por  $1/n$ ).

Para um tenant  $k$ , definimos um vetor de personalização  $e(k) \in \Delta_{n-1}$  (simplex de probabilidade) e um parâmetro de amortecimento  $\kappa \in (0,1)$ .

O *PageRank* personalizado  $S(k) \in \Delta_{n-1}$  resolve:

$$S(k) = (1-\kappa)e(k) + \kappa P S(k).$$

## G.2. Existência e Unicidade

**Teorema G.1 (Perron–Frobenius + Contraction).**

Se  $P$  é estocástica por colunas e primitiva (após *teleport* ela se torna primitiva), então o sistema linear acima tem solução única  $S(k) \in \Delta_{n-1}$ .

*Prova (esboço).*

Rearranjando:  $(I - \kappa P)S(k) = (1-\kappa)e(k)$ .

Como o raio espectral  $\rho(\kappa P) \leq \kappa < 1$ , a matriz  $I - \kappa P$  é inversível. Assim, existe solução única:

$$S(k) = (1-\kappa)(I - \kappa P)^{-1}e(k).$$

Além disso,  $S(k) \in \Delta_{n-1}$  pois é combinação convexa de vetores não-negativos normalizados.  $\square$

### G.3. Convergência de Potência

Considere a iteração:

$$S_{t+1}(k) = (1-\kappa)e(k) + \kappa P S_t(k).$$

**Teorema G.2 (Convergência linear).**

Para qualquer  $S_0(k) \in \Delta^{n-1}$ , a sequência  $\{S_t(k)\}$  converge linearmente para  $S(k)$  com razão  $\kappa$ , i.e.,

$$\|S_t(k) - S(k)\|_1 \leq \kappa^t \|S_0(k) - S(k)\|_1.$$

*Prova.*

A iteração define um mapeamento afim  $F(x) = (1-\kappa)e(k) + \kappa P x$ .

Como  $\|P x - P y\|_1 \leq \|x - y\|_1$  e  $\kappa < 1$ , segue  $\|F(x) - F(y)\|_1 \leq \kappa \|x - y\|_1$ .

Banach  $\Rightarrow$  contração  $\Rightarrow$  convergência única.  $\square$

### G.4. Tratamento de *Dangling Nodes*

Se uma coluna de  $P$  é nula, substituí-la por  $1/n$  mantém estocasticidade e evita absorção.

Alternativamente, pode-se absorver nós pendentes multiplicando o *teleport* diretamente nas colunas nulas:

$$P' = P + D \|e(k)\|_1^{-1} e(k) \mathbf{1}^T, D_{jj} = I[\text{coluna } j \text{ é nula}].$$

### G.5. Sensibilidade a $\kappa$ e $e(k)$

Derivada em relação a  $\kappa$ :

$$\partial \kappa \partial S(k) = -(I - \kappa P)^{-1} e(k) + (I - \kappa P)^{-1} P S(k).$$

Interpretação: aumenta-se  $\kappa \Rightarrow$  mais peso ao grafo, menos ao vetor de personalização.

Sensibilidade a  $e(k)$ :

$$\partial e(k) \partial S(k) = (1-\kappa)(I - \kappa P)^{-1} \mathbf{1}.$$

Norma limitada por  $1-\kappa \|P\|_1 - \kappa$ , com  $\|P\|_1 = 1$ , implicando estabilidade.

### G.6. *Mixing Time* (limite superior)

O tempo para  $\|S_t(k) - S(k)\|_1 \leq \epsilon$  é

$$t \geq \log_k \log(\epsilon / \|S_0 - S\|_1).$$

Como  $\kappa < 1$ ,  $\log_k < 0$  e a cota é finita; tipicamente poucas dezenas de iterações.

## Apêndice H — Verificadores e Guardiões: Lipschitz, Condicionamento e Complexidade

### H.1. Verificador Factual (RAG)

O verificador factual mede a coerência entre resposta e trechos citados.

Definimos uma função de confiança  $C_f \in [0, 1]$  como média ponderada do *score* híbrido (Apêndice B) dos trechos citados.

Se  $S(\cdot)$  é Lipschitz com constante  $LS$ , então:

$$|C_f(x) - C_f(y)| \leq LS \|x - y\|.$$

Isso assegura **estabilidade**: pequenas variações textuais não alteram radicalmente a confiança.

**Complexidade (conceitual):**

Avaliar  $k$  trechos com índice vetorial e BM25 é  $O(k \log N)$  para busca +  $O(k)$  para a agregação.

## H.2. Verificador Numérico

Definimos a aritmética em campo de **ponto fixo** com escala  $\eta$  (p.ex., 100 para centavos). Operações  $+$ ,  $-$ ,  $\times$  são fechadas por arredondamento simétrico até erro  $\leq 2\eta 1$ .

**Invariantes:**

- Conservação de totais:

$$\text{Total} = (\text{Subtotal} - \text{Descontos})(1 + \tau) \pm \epsilon, |\epsilon| \leq \eta c.$$

- Não negatividade de impostos ( $\tau \geq 0$ ).

**Condição de número (condicionamento):**

Para  $\text{Total} = a - b$ ,  $\kappa = |a - b|(|a| + |b|)$ .

Definir políticas que evitem  $|a - b|$  muito pequeno (perigo de cancelamento catastrófico).

## H.3. Verificador Ético (LTL+D)

Checagem de modelo em LTL sobre trilhas finitas tem custo  $O(|\pi| \cdot |\phi|)$  em *runtime* (conceitual).

Ao compor  $m$  políticas  $\phi_1 \wedge \dots \wedge \phi_m$ , a verificação é linear em  $\sum |\phi_i|$ .

**Lipschitzidade por design:** representar sinais discretos (risk, channel, persuasion) com codificação estável reduz jitter lógico.

# Apêndice I — Análise de Sensibilidade do Score Híbrido e Otimização dos Pesos $\alpha, \beta, \gamma, \delta, \zeta$

## I.1. Problema

Dados rótulos de relevância “oráculo”  $y(x, q) \in [0, 1]$  (obtidos por auditoria humana), ajustamos

$$S_w(x, q) = \sum_{i=1}^5 w_i S_i(x, q), \sum_{i=1}^5 w_i = 1, w_i \geq 0$$

para minimizar o erro quadrático:

$$w \in \Delta^4 \min E[(S_w - y)^2].$$

## I.2. Convexidade e KKT

**Proposição.** O problema é convexo (quadrático em  $w$ , domínio simplex).

Condições KKT dão:

$$2 \sum w - 2\mu = \lambda 1 - v, \sum w_i = 1, w_i \geq 0, v_i w_i = 0,$$

onde  $\Sigma_{ij} = E[S_i S_j]$ ,  $\mu_i = E[y S_i]$ .

**Solução fechada** (sem *active set*):

$$w^* \propto \Sigma^{-1} \mu$$

e normaliza-se  $w^*$  no simplex; se algum  $w_i^* < 0$ , aplica-se *projection onto simplex* (método de Michelot).

## I.3. Regularização e Estabilidade

Adicionar  $\lambda \|w\|_2^2$  dá:

$$w^* \propto (\Sigma + \lambda I)^{-1} \mu,$$

reduzindo variância (controle de sobreajuste ao oráculo).

## I.4. Sensibilidade

$$\partial w_i \partial S_w = S_i, \partial w_i \partial w_j \partial^2 S_w = 0.$$

O score é linear nos pesos; portanto, ajuste é estável e interpretável.

# Apêndice J — CMDP Ético: Programação Dinâmica com Restrições e Duais de Lagrange

## J.1. CMDP

Formulamos a política como **CMDP** (MDP com restrições):  
maximize  $E[\sum \gamma r_t]$  sujeito a  $E[\sum \gamma c_t k] \leq dk$  para custos  $c_k$  (éticas/risco).

## J.2. Lagrangiano

$$L(\pi, \lambda) = E[\sum \gamma r_t] - k \sum \lambda_k (E[\sum \gamma c_t k] - dk), \lambda_k \geq 0.$$

O dual é  $\min_{\lambda \geq 0} \max_{\pi} L(\pi, \lambda)$ .

## J.3. Otimalidade e Estabilidade

Sob convexidade/compacidade apropriadas, **forte dualidade** vale e as soluções  $(\pi^*, \lambda^*)$  satisfazem KKT.  
Interpretação:  $\lambda_k$  são “preços éticos” que penalizam violações, estabilizando a política na fronteira **ótima segura**.

# Apêndice K — Controle Afetivo com Saturação: BIBS e Lyapunov para $H_t$ e $P_t$

## K.1. Sistema

$$H_{t+1} P_t = \rho H_t + (1-\rho) \sigma (w^T z_t), | \rho | < 1, = \min\{P^-, \psi(\kappa_1 S_t + \kappa_2 H_t + \kappa_3 C_t)\}.$$

## K.2. BIBS (Bounded-Input Bounded-State)

Se  $|S_t|, |C_t| \leq M$  e  $\sigma, \psi$  são limitadas e Lipschitz, então  $|H_t| \leq 1 - |\rho| (1-\rho) \|\sigma\|^\infty$  (limite geométrico), e  $P_t \leq P^-$ .  
Logo, o sistema é **BIBS estável**.

## K.3. Lyapunov

$$\text{Escolha } V(H) = 21 H^2.$$

$$\Delta V = 21(H_{t+1}^2 - H_t^2) \leq 21(\rho^2 - 1) H_t^2 + \text{termos limitados}.$$

Como  $| \rho | < 1$ , o termo dominante é negativo  $\Rightarrow$  estabilidade.

# Apêndice L — Convergência em Aprendizado Federado com Ruído DP e Comunicação Parcial

## L.1. Setup

Clientes  $k$  executam atualizações locais com **DP-SGD** (gradientes *clipped* + ruído  $N(0, \sigma^2)$ ), depois o servidor agrega com pesos  $n_k/N$ .

## L.2. Taxa de Convergência (conceitual)

Para perdas convexas e Lipschitz, com  $\text{step} \approx 1/t$ :

$$E[F(\theta T)] - F(\theta^*) \leq O(T^{-1}) + O(n\sigma^2),$$

exibindo termo de ruído DP e termo de iteração.

### L.3. Compression/Quantization (conceitual)

Com quantização  $q(\cdot)$  não tendenciosa ( $E[q(g)] = g$ ) e variância limitada, a taxa sofre degradação aditiva  $O(\text{Var}(q))$ , ainda convergente.

## Apêndice M — Limites de Alucinação no RAG: Probabilidade de Factualidade

### M.1. Modelo Binário Conceitual

Seja  $A$  o evento “resposta é factual”. Seja  $R$  o evento “retrieval contém evidência necessária”.

Assumindo geração condicional correta com probabilidade  $p_c$  quando  $R$  ocorre, e alucinação com probabilidade  $p_h$  quando  $R$  não ocorre:

$$\Pr[A] = \Pr[A|R]\Pr[R] + \Pr[A|\neg R]\Pr[\neg R] = p_c r + (1 - p_h)(1 - r),$$

onde  $r = \Pr[R]$  é *recall* do retrieval.

### M.2. Cotas Inferiores

Se  $p_h \approx 1$  (sem evidência tende a errar), então  $\Pr[A] \approx p_c r$ .

Portanto, **maximizar recall** é crucial. Com *critics* e reconsulta, pode-se elevar  $r \rightarrow r'$  e  $p_c \rightarrow p_c'$ , melhorando  $\Pr[A]$ .

## Apêndice N — Robustez a Instruções Adversas: Projeção em Conjunto Seguro e Não-Expansividade

### N.1. Conjunto Seguro

Defina  $S \subset R_m$  como o conjunto de *estados-decisão* permitidos pelas políticas LTL+D e limites de risco/ética/persuasão.

Dado um estado proposto  $y$ , projetamos:

$$\Pi_S(y) = \arg \min_{x \in S} \|x - y\|.$$

### N.2. Não-Expansividade

A projeção em conjunto convexo fechado é **não-expansiva**:

$$\|\Pi_S(y_1) - \Pi_S(y_2)\| \leq \|y_1 - y_2\|.$$

Logo, ruídos/adversidades não amplificam distância de segurança — estabilizando o comportamento da IA.

## Apêndice O — Otimização Multiobjetivo: Economia, Humano e Ética



## O.1. Fronteira de Pareto e Escalarização

Objetivos (Uecon,Uhum,-Uopen\_ethic) geram uma **fronteira de Pareto**.  
Usamos esalarização convexa:

$$\pi \max \alpha U_{econ} + \beta U_{hum} - \gamma U_{open\_ethic}, \alpha, \beta, \gamma \geq 0, \alpha + \beta + \gamma = 1.$$

Sob convexidade, toda solução Pareto-ótima é solução de algum triplo  $(\alpha, \beta, \gamma)$ .

## O.2. Continuidade e Estabilidade

Se objetivos são Lipschitz e convexos, a fronteira é **convexa e fechada**.  
Variações marginais em  $(\alpha, \beta, \gamma)$  produzem variações controladas na política — útil para *A/B de políticas por tenant*.

# Apêndice P — Estabilidade Numérico-Contábil e Reconciliação no ERP (Teoria)

## P.1. Introdução

O objetivo é demonstrar, de forma formal, que as **rotinas contábeis e financeiras conceituais da EAAS** (dupla-entrada, impostos, conciliações, agregações por centros de custos) podem ser modeladas como **sistemas numéricos estáveis e auditáveis**, com **invariantes** que impedem inconsistências acumulativas.

## P.2. Modelo de Dupla-Entrada e Invariantes

Considere um conjunto de contas  $C$  e um conjunto de lançamentos  $L$ .  
Cada lançamento  $\ell \in L$  é um vetor  $v_\ell \in \mathbb{R}^{|C|}$  com soma nula:

$$c \in C \implies \sum_{\ell \in L} v_\ell(c) = 0.$$

A soma de todos os lançamentos no período  $T$  é:

$$V_T = \sum_{\ell \in L_T} v_\ell.$$

**Invariante contábil global:**

$$c \in C \implies \sum_{T \in \mathcal{T}} V_T(c) = 0 \text{ (conservação de valor)}.$$

*Interpretação:* nenhuma sequência de eventos admissíveis cria desequilíbrio contábil.

## P.3. Impostos, Descontos e Linearidade

Seja Subtotal e Disc (desconto). O imposto é  $\tau \geq 0$ . O total:

$$\text{Total} = (\text{Subtotal} - \text{Disc})(1 + \tau).$$

**Proposição P.1 (Linearidade controlada).**

Total é linear em Subtotal e Disc, preservando convexidade e evitando instabilidades por composições não lineares arbitrárias.

*Prova (esboço).*

$(1 + \tau)$  é constante não negativa; logo,  $\text{Total} = (1 + \tau)\text{Subtotal} - (1 + \tau)\text{Disc}$  é função afim.  $\square$

## P.4. Estabilidade Numérica (Moedas e Arredondamento)

Trabalhamos com **ponto fixo** (escala  $\eta$ , p.ex. 100 para 2 casas decimais).

Erro por operação:  $|\epsilon| \leq 2\eta 1$ .

Em uma fatura com  $n$  termos, incerteza cumulativa:

$$|\epsilon_{\text{tot}}| \leq 2\eta n.$$

Definimos tolerância  $\delta_{\text{moeda}}$  (p.ex. 0,01 BRL); exigimos

$$2\eta n \leq \delta_{\text{moeda}}.$$

**Lema P.1 (Bounded Error).**

Sob a política acima, **não** há “explosão” de erro por somas sucessivas.

## P.5. Conciliação Financeira: Igualdade Estrutural

Seja  $\text{Sext}$  a soma de recebimentos confirmados por um terceiro (p.ex., **Stripe**) e  $\text{SERP}$  a soma de recebimentos reconhecidos no ERP.

**Objetivo conceitual:** forçar  $|\text{Sext} - \text{SERP}| \rightarrow 0$ .

**Modelo:** as transações válidas formam um conjunto  $\Omega$ , com emparelhamento  $M \subset \Omega \times \Omega$ .

A função de custo de reconciliação:

$$M_{\min}(i, j) \in M \sum |v_i - v_j| + \lambda U(\text{n\~o casados}),$$

onde  $U$  penaliza elementos sem par.

**Teorema P.1 (Convergência por Idempotência + Penalização).**

Se cada item é processado com **chave idempotente** e o custo  $U$  diverge com o tempo, o sistema conceitual tende a “fechar”  $M$  (casar itens), minimizando  $|\text{Sext} - \text{SERP}|$ .

*Intuição:* cancelar duplicidades e forçar casamento reduz o desbalanceamento assintoticamente.

## P.6. Centros de Custo e Conservação

Para centros  $K$ , o **rateio** define  $\omega_k \geq 0$ ,  $\sum_k \omega_k = 1$ .

Valores alocados  $v_k = \omega_k \cdot v$  mantêm:

$$k \in K \sum v_k = v(\text{conservac\~ao de alocac\~ao}).$$

**Proposição P.2.**

Se todos os lançamentos obedecem dupla-entrada e rateios conservativos, os **balancetes por centro** mantêm coerência com o **balanço global**. □

## P.7. Robustez a Estornos e Reembolsos

Um estorno  $v \sim$  deve satisfazer  $v \sim -v$  para o item original.

Logo:

$$V T' = V T + v + (-v) = V T.$$

**Corolário P.1.**

Estornos “limpos” preservam invariantes e **não alteram** o balanço agregado (somente redistribuem temporalmente).

□

# Apêndice Q — Teoria de Filas no Omnichat: SLA, Prioridades e Estabilidade

## Q.1. Modelo de Fila Multiclasse

Chegadas Poisson  $\lambda_c$  por classe  $c$  (ex.: vendas, suporte, risco).

Tempos de serviço exponenciais com média  $1/\mu_c$ .

Número de atendentes  $m$  (humanos) + “servidor virtual” (IA) com taxa  $\mu_{IA}$  e políticas LTL (Ap. C).

## Q.2. Estabilidade (Regra de Tráfego)

### Teorema Q.1.

O sistema é **estável** se a carga total  $\rho$  satisfaz

$$\rho = c \sum m \mu_c + \chi c \mu_{IA} \lambda c < 1,$$

onde  $\chi c \in [0, 1]$  indica fração de casos classe  $c$  elegíveis à IA sem handoff (compliance).

*Intuição:* a capacidade combinada (humanos + IA em conformidade) deve superar a demanda efetiva.

## Q.3. Prioridades e SLA

Para classes com **prioridade** (p.ex., risco > vendas), use **filas com prioridade não preemptiva**.

Tempo médio de espera aproximado (M/M/1 com prioridade, forma conceitual):

$$E[W_c] \approx (1 - \rho) \mu_{eff, c, \rho}, \mu_{eff, c} = m \mu_c + \chi c \mu_{IA}.$$

**Meta SLA:**  $\Pr(W_c \leq s_c) \geq 1 - \epsilon$ .

Dimensionamos  $m$  e envolvimento IA ( $\chi c$ ) até cumprir a restrição.

## Q.4. Handoff IA → Humano

Suponha probabilidade de handoff  $h_c$  para classe  $c$ .

Evoluímos um **dois estágios**: IA (rápida) + Humano (lento).

Tempo total médio:

$$E[T_c] = E[T_{IA, c}] + h_c E[T_H, c],$$

com  $E[T_{IA, c}] \ll E[T_H, c]$ .

**Otimização:** reduzir  $h_c$  por **melhor grounding** (Ap. B) e **políticas de decisão** (Ap. C), respeitando ética/risco.

# Apêndice R — Preço Dinâmico: Convexidade, Equilíbrio e Regret

## R.1. Setup

Preço  $p$ , demanda  $D(p)$ , receita  $R(p) = p \cdot D(p)$ .

Assuma  $D'(p) < 0$  (demanda decrescente).

Curva de custo  $C(q)$  convexa,  $q = D(p)$ .

## R.2. Otimalidade Estática

Maximizar  $\Pi(p) = R(p) - C(D(p))$ .

**Condição de primeira ordem:**

$$\Pi'(p) = D(p) + p D'(p) - C'(D(p)) \cdot D'(p) = 0.$$

Para demandas lineares  $D(p) = a - bp$ , solução

$$p^* = \frac{2ba + 2C'(a - bp^*)}{2b}.$$

Com  $C'$  aproximadamente constante localmente,  $p^*$  tem forma fechada.

## R.3. Convexidade Local

Se  $D$  é concava e  $C$  convexa,  $\Pi$  é concava  $\Rightarrow$  máximo único.

#### Teorema R.1.

Sob hipóteses de regularidade, o problema de precificação é **concavo** em  $p$ ; logo, existe solução ótima única.

## R.4. Regret Dinâmico (conceitual)

Em ambiente não estacionário, definimos *regret* contra *benchmark* ótico em janelas:

$$\text{Regret}_T = t - 1 \sum_{t=1}^T (\Pi(p_t^*) - \Pi(p_t)).$$

Com atualização suave e feedback da IA (Apêndices B, D), espera-se  $\text{Regret}_T = o(T)$  sob variação total limitada do ambiente.

# Apêndice S — Grafos Produto-Serviço e Atribuição de Recursos: Ótimos e Heurísticas

## S.1. Grafo Bipartido e Matching

Produtos/Serviços  $U$  e Recursos  $V$ .

Custos  $c_{uv}$  (tempo, distância, utilização).

Problema de **matching mínimo**:

$$X \in \{0, 1\}^{|U| \times |V|} \text{ min } u, v \sum_{u,v} c_{uv} X_{uv}, v \sum_{u,v} X_{uv} = 1, u \sum_{u,v} X_{uv} = 1.$$

#### Teorema S.1.

Se  $|U| = |V|$  e  $c$  satisfaz desigualdades do tipo quadrado métrico, o algoritmo Húngaro encontra ótimo global.

## S.2. Recursos Múltiplos e Janelas de Tempo

Para múltiplos recursos e janelas  $[a_i, b_i]$ , o problema vira **VRPTW** (conceitual).

Exigimos heurísticas (inserção, *tabu*, GRASP) com garantia de **factive** e avaliação por *gap* empírico.

## S.3. Robustez e Tie-Breaks Estáveis

Quando  $c_{uv}$  é quase empatado, defina **tie-breaks** determinísticos (ex.: prioridade por satisfação histórica), preservando **consistência temporal** e evitando alternância caótica.

# Apêndice T — Auditoria, Não-Repúdio e Integridade de Logs

## T.1. Estrutura de Log

Cada evento  $e$ :

$$e = (\text{tenant}, \text{actor}, \text{tool}, \text{request}, \text{response}, \text{status}, \text{timestamp}, \text{signature}).$$

**Assinatura/HMAC** garante **não-repúdio** (o emissor não pode negar o envio).

## T.2. Integridade e Cadeia Imutável

Criamos um **hash-chain**:

$$h_0 = H(e_0), h_{i+1} = H(e_{i+1} || h_i),$$

onde  $H$  é hash criptográfico.  
Qualquer adulteração rompe a cadeia  $\Rightarrow$  auditor detecta.

### T.3. Propriedade de Completude

#### Teorema T.1.

Se todo evento é assinado e encadeado, a probabilidade de remoção indetectável de um evento é **negligenciável** (na hipótese de resistência a colisões de  $H$ ).

*Interpretação:* logs são **forensic-ready**.

### T.4. Reconciliação de Logs Multi-Fonte

Logs de IA, Gateway, ERP, Pagamentos formam visões  $\{L_i\}$ .

Defina o **pullback** (interseção temporal/semântica)  $P = \bigcap_i L_i$ .

#### Proposição T.1.

Se cada  $L_i$  é íntegro e sincronizado, então  $P$  é **suficiente** para reconstituir a história canônica dos eventos (até relógios com *skew* limitado).

## Apêndice U — Robustez a *Domain Drift* na Knowledge Base: Testes, Métricas e Políticas de Atualização

### U.1. Visão Geral

O *domain drift* descreve mudanças estatísticas no conteúdo/semântica que abastece a KB (docs internos, sites autorizados, regulatórios, catálogos). Para manter a IA **factual**, **estável** e **coerente**, precisamos:

1. **Detectar** drift (mudanças nas distribuições textuais/semânticas);
2. **Quantificar** magnitude e direção;
3. **Decidir** políticas de reindexação, *downweighting* e *roll-back*.

### U.2. Métricas de Deslocamento de Distribuições

Sejam duas amostras de chunks vetoriais de períodos diferentes:  $X = \{x_i\}_{i=1}^n$  e  $Y = \{y_j\}_{j=1}^m$ , com embeddings  $x_i, y_j \in \mathbb{R}^d$ .

#### U.2.1. PSI (Population Stability Index)

Para *bins*  $b$ , com proporções  $p_b$  (baseline) e  $q_b$  (atual):

$$\text{PSI} = b \sum (q_b - p_b) \ln p_b q_b.$$

**Interpretação:**  $\text{PSI} \approx 0$  indica estabilidade;  $\text{PSI} > 0.25$  (regra comum) sugere drift relevante.

#### U.2.2. Divergência de Kullback–Leibler (KL)

Se  $P$  e  $Q$  são densidades (estimadas por KDE) no espaço vetorial:

$$\text{DKL}(Q \| P) = \int Q(z) \ln P(z) Q(z) dz.$$

**Propriedade:**  $\text{DKL} \geq 0$  e  $= 0$  sse  $P = Q$  quase certamente.

#### U.2.3. MMD (Maximum Mean Discrepancy)

Para kernel positivo-definido  $k$ :

$$\text{MMD2}(X,Y)=n^{-1}\sum_{i,i'}k(x_i,x_{i'})+m^{-1}\sum_{j,j'}k(y_j,y_{j'})-nm^{-1}\sum_{i,j}k(x_i,y_j).$$

**Teste:** rejeitar  $H_0:P=Q$  se  $\text{MMD2}$  excede limiar (obtido por *permutation*).

#### U.2.4. Distância de Wasserstein (Earth Mover's)

Para distribuições empíricas  $P_X, P_Y$ :

$$W_1(P_X, P_Y) = \inf_{\pi \in \Pi(P_X, P_Y)} \int E(x, y) d\pi \sim \int \|x - y\| d\pi.$$

**Intuição:** custo mínimo de “transportar massa” de  $X$  para  $Y$ .

### U.3. Detecção Sequencial e *Change-Points*

#### U.3.1. CUSUM Conceitual

Para estatística  $Z_t$  (p.ex., *score* de similaridade média), definimos:

$$S_t = \max\{0, S_{t-1} + (Z_t - v)\}, S_0 = 0,$$

sinalizando mudança quando  $S_t > h$ .

$v$  é média esperada sob  $H_0$ ;  $h$  controla probabilidade de falso alarme.

#### U.3.2. Teste de Page–Hinkley (conceitual)

Monitora variação acumulada da média, com gatilho quando desvio excede limiar.

### U.4. Política de Atualização e *Downweighting*

Se  $\text{PSI}$  ou  $\text{DKL}$  excedem limites, aplicamos:

- **Downweighting** de documentos antigos conflitantes;
- **Reindexação focada** (refrescar clusters/temas com maior drift);
- **Janela deslizante** temporal para deprecicar conteúdos obsoletos;
- **Gate humano** para tópicos críticos (compliance).

**Função de frescor adaptativa** (extensão):

$$\text{Sfresco}(x) = \exp[-\lambda(t) \cdot \text{idade}(x)], \lambda(t) = \lambda_0 + \lambda_1 \cdot \text{DriftIndex}(t),$$

aumentando o decaimento quando o drift global cresce.

### U.5. Estabilidade e Garantias

**Teorema U.1 (Estabilidade por *Bounded Drift*).**

Se os indicadores de drift obedecem  $\sup_t \text{MMD}(X_t, X_{t-1}) \leq \Delta < \infty$  e a política de *downweighting* é Lipschitz em  $\Delta$ , então o *score* de RAG híbrido (Ap. B) permanece **Lipschitz** no tempo:

$$|S_t - S_{t-1}| \leq L \cdot \Delta.$$

*Prova (esboço):* composição de funções Lipschitz (kernels, pesos) preserva propriedade; deriva do teorema de composição.  $\square$

## Apêndice V — Modelagem de Risco, Chargeback e Decisão Custo-Sensível

## V.1. Probabilidade de Fraude e Decisão Ótima

Seja  $X$  o vetor de atributos da transação, e  $Y \in \{0, 1\}$  indicador de fraude.

Estimamos  $p(x) = \Pr(Y=1|X=x)$ .

Definimos utilidade esperada ao **aprovar**:

$U_{\text{approve}}(x) = \text{margem esperada} \cdot (1 - p(x)) - \text{perda chargeback} \cdot p(x)$ .

Decisão ótima conceitual:

$\text{aprovar} \Leftrightarrow U_{\text{approve}}(x) \geq 0 \Leftrightarrow p(x) \leq \pi + L\pi$ .

**Limiar ótimo**  $\tau^* = \pi + L\pi$  depende dos **custos** do negócio.

## V.2. Curvas ROC e Operating Point

Ajustamos o **ponto de operação** na curva ROC de acordo com  $\tau^*$ .

**Custo esperado:**

$C(\tau) = L \cdot \text{FNR}(\tau) \cdot P + \pi \cdot \text{FPR}(\tau) \cdot (1 - P)$  (exemplo de custo assimétrico).

$P = \Pr(Y=1)$  é base rate de fraude.

## V.3. Cadeia de Markov de Disputas

Estados  $\{S_0 = \text{limpo}, S_1 = \text{disputa}, S_2 = \text{perdido}, S_3 = \text{ganho}\}$ .

Matriz de transição  $T$ .

Perda esperada por transação:

$E[L] = \pi_0 T_k \cdot \ell$ ,

onde  $\ell$  dá custos por estado terminal.

## V.4. Stress Tests e Worst-Case

Modela-se choque de base rate  $P \rightarrow P' = P + \delta$ .

Garantia: estratégia  $\tau^*$  é **minimax** para  $\delta$  em intervalo compacto se a função de custo é convexa em  $p$ .

*Intuição:* definimos *guard bands* para operar com **margem de segurança**.

# Apêndice W — Do KPI ao Formalismo de Políticas: Mapeamento para CMDP e LTL+D

## W.1. KPIs de Negócio $\rightarrow$ Restrições Formais

Exemplos:

- **CSAT**  $\geq s_{\min}$ ,
- **FCR**  $\geq f_{\min}$ ,
- **Tempo médio**  $\leq t_{\max}$ ,
- **Alçada ética** (persuasão  $\leq P^-$ , PII sem consentimento = 0).

Mapeamos para CMDP (Ap. J) com custos  $ck$ :

$E[\sum \gamma t c_1] \leq d_1$  (viol. ética),  $E[\sum \gamma t c_2] \leq d_2$  (viol. SLA), ...

e simultaneamente para LTL:

$G(\text{answer} \rightarrow \text{Ocitation}), G(\text{risk} > \tau \rightarrow \text{Ohandoff})$ .

## W.2. Dualidade e Interpretação

Multiplicadores de Lagrange  $\lambda_k$  (Ap. J) refletem **preço sombra** das restrições.

A política ótima resolve:

$$\pi_{\max} E[\sum y_{trt}] - k \sum \lambda_k (E[\sum y_{tctk}] - dk).$$

**Sensibilidade:** variações marginais nos *targets*  $dk$  alteram  $\lambda_k$  e, portanto, o comportamento da IA (mais/menos conservadora).

## W.3. Consistência KPI $\leftrightarrow$ LTL

**Teorema W.1 (Compatibilidade).**

Se as restrições CMDP são factíveis e as fórmulas LTL não são mutuamente excludentes, então existe uma política  $\pi$  tal que as metas de KPI são atingíveis e as fórmulas LTL são satisfeitas (em sentido de trilhas admissíveis).

*Prova (esboço):* interseção não vazia de conjuntos factíveis; aplicar teorema de separação convexa.  $\square$

# Apêndice X — Teoria de Observabilidade: Identificabilidade, Estimadores e Testes Sequenciais

## X.1. Identificabilidade de Métricas

Sistemas de telemetria (latência, custo, *tool-usage*) devem permitir reconstrução de métricas de interesse.

**Definição (Identificabilidade):** um parâmetro  $\theta$  é identificável se distribuições distintas de  $\theta$  geram distribuições distintas dos observáveis  $Y$ .

**Lema X.1:** Se o mapa  $\theta \mapsto P_\theta(Y)$  é injetivo em classe  $\Theta$ , então  $\theta$  é identificável.

## X.2. Estimadores e ICs

Para métrica  $\mu = E[Y]$ , estimador  $\mu^\wedge = n^{-1} \sum Y_i$ .

**IC (normal assintótico):**

$$\mu^\wedge \pm z_{1-\alpha/2} n \sigma^\wedge.$$

Para proporções (ex.: FCR), usar **Clopper–Pearson** (exato) ou **Wilson** (aprox).

## X.3. Propagação de Erro

Se  $Z = g(X_1, \dots, X_k)$  suave,

$$\text{Var}(Z) \approx \nabla g^T \Sigma \nabla g,$$

com  $\Sigma$  matriz de covariâncias de  $X$ .

**Uso:** compor latência total de múltiplos módulos.

## X.4. A/B Testing, Alpha Spending e Bandits

### X.4.1. A/B Clássico

$$\text{Estatística } Z = \frac{\sigma^A^2/n_A + \sigma^B^2/n_B}{\mu^A - \mu^B}.$$

Controle de erro tipo I  $\alpha$ ; poder  $1 - \beta$ ; tamanho amostral por fórmula clássica.



### X.4.2. Sequential Testing (alpha spending)

Passos interinos com *alpha spending* at tal que  $\sum \alpha_t \leq \alpha$ .

**Vantagem:** parar cedo com evidência suficiente.

### X.4.3. Multi-Armed Bandits (Conceitual)

Exploração vs. exploração: UCB/Thompson Sampling brindam *regret* sublinear  $o(T)$  sob hipóteses padrão.

Aplicação: variações de políticas de IA (pesos  $\alpha, \beta, \gamma$ , Ap. O) por tenant.

## X.5. Power e Minimum Detectable Effect (MDE)

Para diferença de médias  $\Delta$ , desvio  $\sigma$  e  $\alpha, \beta$  dados:

$$n \approx 2(\Delta/\sigma z_{1-\alpha/2} + z_{1-\beta})^2.$$

**Uso:** dimensionar experimentos de políticas (persuasão, regras, prompts conceituais) com **poder adequado**.

# Apêndice Y — Harmonização Multilíngue e Equivalência Semântica (pt-BR / EN / ES)

## Y.1. Motivação

A IA da EAAS deve operar com **coerência semântica** entre idiomas (pt-BR, inglês, espanhol) em:

(i) respostas; (ii) citações RAG; (iii) políticas LTL; (iv) métricas CRM/ERP.

Problema: **equivalência de conteúdo** sob variações léxicas, sintáticas e pragmáticas.

## Y.2. Espaço Semântico Multilíngue

Sejam embeddings multilíngues  $\phi_\ell: X_\ell \rightarrow \mathbb{R}^d$ , para idioma  $\ell \in \{\text{pt}, \text{en}, \text{es}\}$ .

Exigimos **isometria aproximada**:

$$\forall x, \|\phi_{\text{pt}}(x) - \phi_{\text{en}}(\text{Tr}_{\text{pt} \rightarrow \text{en}}(x))\| \leq \epsilon, \|\phi_{\text{pt}}(x) - \phi_{\text{es}}(\text{Tr}_{\text{pt} \rightarrow \text{es}}(x))\| \leq \epsilon.$$

Aqui  $\text{Tr}_\ell \rightarrow \ell'$  denota a tradução conceitual.

**Lema Y.1 (Lipschitzidade de Tradução).**

Se  $\phi_\ell$  e  $\text{Tr}_\ell \rightarrow \ell'$  são Lipschitz com constantes  $L_\phi, L_T$ , então:

$$\|\phi_\ell(x) - \phi_{\ell'}(\text{Tr}_\ell \rightarrow \ell'(x))\| \leq L_\phi L_T \|x\|.$$

Interpretação: estabilidade de equivalência cresce com a qualidade dos mapeamentos.

## Y.3. Equivalência de Conteúdo e “Fidelidade”

Definimos uma medida de fidelidade  $F \in [0, 1]$  entre pares de frases/documentos:

$$F(x_\ell, y_{\ell'}) = \exp(-\alpha \|\phi_\ell(x_\ell) - \phi_{\ell'}(y_{\ell'})\|) \cdot \text{consistência factual}_J(\text{NER}, \text{units}, \text{dates})$$

onde  $J \in [0, 1]$  penaliza divergências de entidades/unidades/datas.

**Teorema Y.1 (Cota inferior de fidelidade).**

Se os desvios de NER/unidades/datas são nulos e o erro de isometria é  $\leq \epsilon$ , então  $F \geq e - \alpha \epsilon$ .

## Y.4. Política Multilíngue na LTL

Fórmulas LTL são **idioma-agnósticas**: os predicados (`consent`, `collect_PII`) são definidos no **esquema lógico** e não na linguagem natural.

A camada de linguagem **projeta** intenções em predicados canônicos antes da verificação (Ap. C).

# Apêndice Z — Grounded Persuasion: Persuasão Limitada com Provas de Segurança

## Z.1. Variáveis e Limites

Retomemos  $S_t$  (sentimento),  $H_t$  (humor),  $C_t$  (contexto),  $P_t$  (persuasão).  
Definimos **limite global**  $P^-$  (tenant) e **limites por canal**  $p_{WA}, p_{web}, p_{email}$ .

$$P_t = \min\{P^-, \psi(\kappa_1 S_t + \kappa_2 H_t + \kappa_3 C_t)\}.$$

## Z.2. Restrições Deonticas

$G((channel=WA \wedge P_t > p_{WA}) \rightarrow Fob \text{ execute\_action}), G((channel=web \wedge P_t > p_{web}) \rightarrow Fob \text{ execute\_action}).$

Estas asseguram **bounded persuasion** por canal (Ap. C).

## Z.3. Propriedade de Segurança

**Teorema Z.1 (Bounded Persuasion Safety).**

Se  $\psi$  é Lipschitz e  $S, H, C$  são limitados, então  $P_t \leq \max(P^-, p_{canal})$ .

Com LTL, sempre que  $P_t > p_{canal}$ , a execução é **proibida**. Logo, **nenhuma ação** persuasiva acima do limite é tomada.

*Prova (esboço):* segue da definição de  $P_t$  e dos guardas LTL.  $\square$

## Z.4. Métrica de Wellbeing

Para quantificar “pressão” exercida:

$$W_t = 1 - P_t,$$

com  $W_t \in [0, 1]$ .

Políticas podem impor  $G(W_t \geq \omega_{min})$  garantindo **bem-estar mínimo** na interação.

# Apêndice AA — Planejamento em Árvores (ToT/GoT) como Árvores de Prova e Limites de Profundidade

## AA.1. Estrutura de Planejamento

Representamos um plano como árvore  $T=(N,E)$  com nós  $n=(s,a)$  e custo  $c(n)$ .

O objetivo é encontrar caminho raiz  $\rightarrow$  folha com **max-score**:

$$score(n) = \lambda_1 Q^A(n) - \lambda_2 risk(n) + \lambda_3 explain(n).$$

## AA.2. Árvores de Prova

Cada nó contém **obrigações**  $\Phi$  (submetas ou políticas) e **evidências**  $E$  (citações RAG).

Uma folha é **válida** se  $\Phi \subseteq E$  (todas as exigências cobertas).

**Lema AA.1 (Somatório de Evidências).**

Se cada aresta adiciona evidência disjunta, então  $|E_{folha}| = \sum \text{evidência nas arestas}$  (sem sobreposição), simplificando auditoria.

## AA.3. Limites de Profundidade

**Teorema AA.1 (Cut-off de Profundidade).**

Se o custo esperado por profundidade cresce geometricamente  $c_d = c_0 \alpha^d$ ,  $\alpha > 1$ , e o ganho marginal de score decresce geometricamente  $\Delta s_d = s_0 \beta^d$ ,  $\beta < 1$ , então existe  $d^*$  finito tal que  $\forall d > d^*$ ,  $\Delta s_d < c_d$  — logo, **planejamento profundo adicional** é ineficiente.

*Intuição:* garantia conceitual de **profundidade finita ótima**. □

**AA.4. Poda Segura**

Regras de poda preservam **completude relativa**:

pode-se eliminar ramos cujo *upper bound* de score < melhor *lower bound* corrente (branch-and-bound conceitual).

## Apêndice AB — Robustez à Ambiguidade Numérica: Espaço Métrico para Unidades/Moedas/ Datas e Projeções Canônicas

**AB.1. Espaço de Representações**

Seja  $U$  o conjunto de unidades (km, mi, kg, lb),  $M$  de moedas (BRL, USD, EUR),  $D$  de formatos de data. Definimos métricas:

$$d_U(u_1, u_2) = |\log(\kappa(u_1 \rightarrow u_2))|, \quad d_M(m_1, m_2) = |\log(\xi(m_1 \rightarrow m_2))|, \quad d_D(d_1, d_2) = I[\text{formatos distintos}] + \epsilon \cdot |\text{epoch}(d_1) - \text{epoch}(d_2)|.$$

$\kappa, \xi$  são fatores de conversão conceituais (positivos),  $\epsilon > 0$  pequeno.

**AB.2. Projeções Canônicas**

Dada entrada ambígua  $x$ , projetamos em um **representante canônico**:

$$\Pi_U(u) = \arg v \in U_{\text{canon}} \min d_U(u, v), \quad \Pi_M(m) = \arg v \in M_{\text{canon}} \min d_M(m, v), \quad \Pi_D(d) = \arg v \in D_{\text{canon}} \min d_D(d, v).$$

**Propriedade:** projeções em conjuntos finitos são **não-expansivas** por mínimo.

**AB.3. Verificador Numérico Estendido**

Após projeção, todas as operações (soma, produto, imposto) são feitas no **sistema canônico**, reduzindo risco de erro por ambiguidade.

## Apêndice AC — Calibração Probabilística: Brier, LogLoss e Garantias de Decisão

**AC.1. Motivação**

Decisões da IA (aprovar compra, propor upgrade, escalar humano) dependem de **probabilidades** calibradas. Probabilidades não calibradas  $\Rightarrow$  limiares subótimos (Ap. V).

**AC.2. Métricas de Calibração**

- **Brier Score** (para classe binária):

$$\text{Brier} = \frac{1}{n} \sum_{i=1}^n (p_i - y_i)^2.$$

- **LogLoss**:

$$\text{LogLoss} = -n \sum_{i=1}^n [y_i \log p_i + (1 - y_i) \log (1 - p_i)].$$

### AC.3. Curvas de Confiabilidade

Particiona-se  $[0, 1]$  em *bins*; compara-se **média de p** vs. **frequência real**.  
 Erro de calibração por *bin*  $b$ :

$$E_b = |p_b - y_b|.$$

**ECE** (Expected Calibration Error):

$$\text{ECE} = \sum_b \omega_b E_b, \sum_b \omega_b = 1.$$

### AC.4. Garantias de Decisão (conceitual)

Se a calibração produz  $\sup_b E_b \leq \epsilon$ , então a decisão custo-sensível com limiar  $\tau^*$  (Ap. V) tem **erro de decisão** limitado por função crescente em  $\epsilon$ .

Intuição: **pequeno erro de calibração**  $\Rightarrow$  **pequena degradação de utilidade**.

## Apêndice AD — Auditoria Causal por Contrafactuais: Teoria, Testes e Garantias

### AD.1. Motivação e Quadro Geral

Para que a IA da EAAS seja **explicável** e **auditável**, precisamos responder: “O que teria acontecido se uma decisão alternativa fosse tomada?”

Usamos o arcabouço de **Causal Inference** com *Structural Causal Models* (SCM) no sentido de Pearl.

#### AD.1.1. SCM Conceitual

Sejam variáveis endógenas  $V = \{V_1, \dots, V_n\}$  e exógenas  $U = \{U_1, \dots, U_m\}$ .

Cada  $V_i$  é dado por:

$$V_i := f_i(\text{Pa}(V_i), U_i),$$

onde  $\text{Pa}(V_i)$  são os pais de  $V_i$  no **DAG causal**  $G$ .

Um **do-intervention** fixa uma variável:  $\text{do}(X=x)$ , removendo as arestas que entram em  $X$ .

#### AD.1.2. Três Níveis (Associação, Intervenção, Contrafactual)

1. **Observacional**:  $P(Y|X)$
2. **Intervencional**:  $P(Y|\text{do}(X=x))$
3. **Contrafactual**:  $P(Y_x|X=x', Y=y')$

A auditoria causal que visamos é **conceitual**: **inferir tendências e plausibilidades** sem executar intervenções reais no ambiente do cliente.

### AD.2. Identificação Intervencional (Back-Door, Front-Door)

#### AD.2.1. Back-Door

Se  $Z$  bloqueia todos os caminhos *back-door* de  $X$  a  $Y$ , então:

$$P(Y|\text{do}(X=x)) = \sum_z P(Y|X=x, Z=z)P(Z=z).$$

*Interpretação*: ajustar por confundidores  $Z$  remove viés de seleção.

#### AD.2.2. Front-Door

Se  $Z$  satisfaz as condições *front-door*:

$$P(Y|\text{do}(X=x)) = \sum_z P(z|X=x) \sum_{x'} P(Y|z, X=x') P(X=x').$$

*Intuição:* mediação observada por Z permite identificar efeito de X em Y mesmo com confundidor não observado.

### AD.3. Contrafactuais (Nível 3)

Dado resultado observado  $Y=y'$  sob  $X=x'$ , o **contrafactual**  $Y_x$  responde: “Qual seria Y se X tivesse sido  $x$ ?”.  
O cálculo conceitual segue:

1. **Abdução:** inferir U consistente com observação.
2. **Ação:** fixar  $\text{do}(X=x)$ .
3. **Predição:** recomputar Y.

#### Propriedade AD.1 (Coerência de Contrafactuais).

Se o SCM é **Markoviano** e **não-cíclico**, contrafactuais são bem-definidos e consistentes com as distribuições intervencionais.

### AD.4. Auditoria Causal em Decisões de IA

Considere decisão binária  $A \in \{0, 1\}$  (ex.: *aprovar reembolso*).

Outcome Y (ex.: *satisfação, custo, fraude*).

Queremos o **ACE** (Average Causal Effect):

$$\text{ACE} = E[Y|\text{do}(A=1)] - E[Y|\text{do}(A=0)].$$

#### Teste de Robustez:

Se ACE troca de sinal sob pequenas perturbações de suposições (confundidores não medidos), a política é **causalmente frágil**.

Adotamos **sensibilidade causal** (Rosenbaum) conceitual: varrer *odds ratios* de confundimento até um  $\Gamma$  que muda a conclusão — política só é “robusta” se  $\Gamma$  for alto.

### AD.5. Garantias e Interpretação

#### Teorema AD.1 (Coerência Causal–LTL).

Se a política impõe LTL que força handoff em risco alto, então qualquer **decisão automática** entra em zona onde o DAG causal exclui efeitos adversos não mitigados.

*Esboço:* regiões de ação automática são recortes do estado onde predicados “risco” e “ética” estão abaixo de limiares; assim, os caminhos causais de dano são **diluídos** por design. □

## Apêndice AE — *Bilevel Learning* no RAG Híbrido: Pesos Globais (Meta) e Seleção de Evidências (Base)

### AE.1. Formulação Conceitual

Camada superior escolhe pesos  $w=(\alpha, \beta, \gamma, \delta, \zeta)$  (Ap. B).

Camada inferior seleciona evidências  $E \subseteq D$  maximizando fidelidade e cobertura.

$$w \in \Delta \min E[L(S_w(E(q)), y(q))] \text{ s.t. } E(q) = \arg E \subseteq D \max \text{cover}(E, q) - \lambda \text{cost}(E).$$

#### AE.1.1. Observações

- **Upper level (meta)** resolve pesos;
- **Lower level (base)** escolhe conjunto de trechos (knapsack/constrained selection).

## AE.2. Propriedades e Soluções Conceituais

Se  $L(\cdot)$  é convexa em  $Sw$  e  $Sw$  linear em  $w$ , então **nível superior é convexo**.

O nível inferior é um **subproblema combinatório**; adotamos **relaxações** contínuas (por exemplo, variáveis  $x_i \in [0, 1]$  que indicam “fração” de evidência).

**KKT meta:**

$\nabla w E[L] + \sum_j \lambda_j \nabla w g_j(w) = 0$ , com  $g_j$  restrições do simplex.

**Conceito:** aprendemos  $w$  que favorecem evidências que maximizam fidelidade/autoridade para cada *tenant*.

## AE.3. Estabilidade

**Teorema AE.1 (Estabilidade Meta).**

Se  $L$  é Lipschitz e  $E(q)$  varia de forma semicontínua (relaxação convexa), então o problema bilevel tem **solução estável** ao ruído de consulta.

*Esboço:* composição Lipschitz + ótimos de problemas convexos com mapeamentos semicontínuos.  $\square$

# Apêndice AF — *Cost Throttling* da IA: Orçamentos, SLO de Custo e Garantias

## AF.1. Meta

A IA deve manter **custo por interação** sob um teto definido por tenant, sem degradar SLO de qualidade/latência.

## AF.2. Formalismo

Defina custo  $K = \sum tct$  por conversa; meta  $K \leq B$  (orçamento).

Definimos Lagrangiano:

$L = E[\text{Qualidade}] - \lambda(E[K] - B) - \mu(p95\_lat - \tau)$ .

Otimizamos com  $\lambda, \mu \geq 0$ .

**Interpretação:**  $\lambda$  “preço do custo”;  $\mu$  “preço da latência”.

## AF.3. Garantias

**Teorema AF.1.**

Sob convexidade/regularidade, existe ponto de sela  $(\pi^*, \lambda^*, \mu^*)$  tal que a política  $\pi^*$  cumpre orçamentos e SLOs esperados.

*Esboço:* dualidade forte em problemas convexos com restrições.  $\square$

## AF.4. Estratégias

- **Depth control** em cadeias de reflexão (capar *loops* quando *gain* marginal < custo marginal);
- **RAG adaptativo** (menos documentos quando fidelidade já é alta);
- **Chamada a ferramentas sob demanda** (apenas quando critério crítico aciona).

# Apêndice AG — Saturação Cognitiva e Anti-Overlooping: Critérios de Parada Ótimos

## AG.1. Problema

Cadeias de “planejar–criticar–refletir” podem **demorar** e **custar**. Queremos **parar** quando retorno marginal não justifica custo adicional.

## AG.2. Regra de Parada

Seja  $G_d$  o ganho esperado no ciclo  $d$  e  $C_d$  o custo (tokens/latência).  
Parar quando:

$$G_d - C_d \leq \epsilon \text{ ou } C_d G_d \leq \tau.$$

**Teorema AG.1 (Ótimo de Myopically Optimal Stopping).**

Se  $G_d$  é decrescente e  $C_d$  não-decrescente, existe um  $d^*$  finito que maximiza valor líquido.

*Prova:* segue de propriedades de sequências unimodais e critério de razão.  $\square$

## AG.3. Estabilidade (Lyapunov de Loop)

Defina  $V(d) = \sum_{i=1}^d (C_i - G_i)$ .

Se  $\Delta V(d) = V(d+1) - V(d) = C_{d+1} - G_{d+1} \geq \epsilon > 0$  após certo  $d_0$ , então  $V$  cresce  $\Rightarrow$  parar em  $d_0$ .

*Intuição:* evita “overlooping” com critério de energia.

# Apêndice AH — *Knowledge Poisoning*: Modelo Estatístico, Filtros de Robustez e Detecção

## AH.1. Ameaça

Fontes externas maliciosas podem inserir **conteúdo envenenado** na KB (ex.: instruções sutis que distorcem respostas futuras).

## AH.2. Modelo Conceitual

Classifique documentos com *score* de confiabilidade  $A(x) \in [0, 1]$  e vetor semântico  $v_x$ .

Define-se **poison score**:

$$\Pi(x) = \lambda_1 \cdot \text{Outlier}(v_x) + \lambda_2 \cdot (1 - A(x)) + \lambda_3 \cdot \text{Contrad}(x, D),$$

onde:

- **Outlier**( $v_x$ ): distância de Mahalanobis a cluster confiável;
- **Contrad**: grau de contradição lógica/semântica com base consolidada.

## AH.3. Testes Estatísticos

- **Mahalanobis**:  $M^2 = (v_x - \mu)^T \Sigma^{-1} (v_x - \mu)$  com *p-value* por  $\chi^2$ ;
- **Entailment-contradiction** conceitual: medir  $p(\text{contrad} | x, \text{corpus})$  em escala  $[0, 1]$ .

## AH.4. Políticas de Gate

- **Quarentena** se  $\Pi(x) > \tau_P$ ;
- **Downweight** se  $\tau_{low} < \Pi(x) \leq \tau_P$ ;
- **Manual QC** para itens em zonas cinzentas.

**Teorema AH.1 (Redução de Risco por Thresholding).**

Sob hipóteses de separabilidade estatística (distribuições *benignas* vs. *maliciosas* com sobreposição limitada), existe  $\tau_P$  que maximiza  $1 - \text{BEP}$  (Balanced Error Probability).  $\square$

## AH.5. Estabilidade

Se supt taxa de *poison* aceita é  $\leq \epsilon$ , e o RAG híbrido rebaixa autoridade de itens com alto  $\Pi(x)$ , então a probabilidade de alucinação maliciosa  $\downarrow O(\epsilon)$ .

*Intuição:* a cadeia de confiança com LTL (citação obrigatória) mitiga exploração.

# Apêndice AI — Multi-Objective Reinforcement Learning: Otimalidade de Pareto e Conformidade Ética

## AI.1. Fundamentos

O agente IA da EAAS opera em múltiplas dimensões de recompensa:

$rt = (r_{tecon}, r_{thum}, r_{teth}, r_{ttech})$ ,

representando ganhos econômicos, humanos (satisfação), éticos e técnicos.

Queremos políticas  $\pi$  que sejam **Pareto-ótimas**:

$\nexists \pi' : R(k)(\pi') \geq R(k)(\pi), \forall k$ , com estrita desigualdade em algum  $k$ .

## AI.2. Fronteira de Pareto

A fronteira  $P$  é o conjunto de pontos não-dominados:

$P = \{R(\pi) \mid \nexists \pi' : R(\pi') > R(\pi) \text{ em todas as dimensões}\}$ .

Visualmente, representa a “curva eficiente” de políticas que equilibram lucro, ética e empatia.

### Teorema AI.1 (Existência).

Se o espaço de políticas  $\Pi$  é compacto e  $R(\pi)$  contínua, então  $P \neq \emptyset$ .

*\*Prova resumida:* segue do Teorema de Weierstrass; o máximo existe em subconjunto convexo compacto.

## AI.3. Escalarização Paramétrica

Usamos pesos  $\lambda = (\lambda_1, \dots, \lambda_K)$  com  $\sum \lambda_k = 1$ :

$J(\pi; \lambda) = \sum \lambda_k R(k)(\pi)$ ,

permitindo selecionar pontos específicos na fronteira.

Para ética:

$\lambda_{eth} \geq \lambda_{econ}/2$ ,

garantindo que lucro não sobrepuje moralidade.

## AI.4. Regularização Ética

Introduzimos penalidade:

$L_{eth} = \mu t \sum I[\text{violac\~ao LTL em } t]$ ,

com  $\mu$  crescente se violação persistir.

**Garantia:** política converge para região de “mínima violação” sob  $\mu \rightarrow \infty$ .

# Apêndice AJ — Teoria do *Handoff* Ótimo IA $\rightarrow$ Humano



## AJ.1. Formulação Sequencial

Cada interação é processo de decisão estocástico com probabilidade  $h_t$  de **handoff** no tempo  $t$ .  
Custo esperado:

$$E[C] = t \sum (c_A(t)(1-h_t) + c_H(t)h_t + \phi | \text{erro sem handoff} |).$$

$\phi$  é custo de erro ético.

## AJ.2. Condição de Handoff Ótimo

Handoff ocorre se:

$$E[c_A] - E[c_H] \geq \phi \cdot \text{perro},$$

onde  $\text{perro}$  é probabilidade prevista de resposta incorreta.

**Lema AJ.1 (Bounded Risk Transfer).**

Se  $c_H > \phi \cdot \text{perro}$  sempre, IA não escalará indevidamente; estabilidade garantida.

## AJ.3. Interpretação

Conceitualmente, o sistema transfere controle apenas quando **benefício cognitivo humano** supera **custo de latência e esforço**, mantendo eficiência e segurança.

# Apêndice AK — Robustez de Grafos Semânticos sob Ruído e Percolação

## AK.1. Modelo

O grafo semântico  $G=(V,E)$  sofre ruído em arestas: cada  $e \in E$  é removida com prob.  $p$ .  
Queremos estabilidade do PageRank (Ap. B).

## AK.2. Limiar de Percolação

**Teorema AK.1 (Limiar de Conectividade).**

Para grafo aleatório  $G(n,p)$ , o limiar de conectividade é  $p_c \approx \ln n / n$ .

Se  $p > p_c$ ,  $G$  conectado w.h.p.  $\Rightarrow$  IA mantém coerência semântica global.

## AK.3. Estabilidade do PageRank

Perturbação  $\Delta P$  em matriz de transição  $P$ :

$$\|r' - r\|_1 \leq (1 - \alpha) \|\Delta P\|_1,$$

com  $\alpha$  fator de amortecimento.

Logo, ruídos pequenos em arestas  $\rightarrow$  desvios controlados em ranking.

# Apêndice AL — *Explainability Budgets* e Quantificação de Transparência

## AL.1. Motivação

Toda decisão autônoma precisa gerar explicação proporcional à complexidade do raciocínio.

## AL.2. Modelo de Orçamento

Cada explicação custa  $E_t$  (em tempo ou recursos).  
Orçamento total  $BE$ .  
Queremos maximizar **entendimento médio** UE:

$$\max_t \sum u(E_t) \text{ s.a. } \sum E_t \leq BE.$$

$u(\cdot)$  é função de utilidade crescente côncava (diminishing returns).

**Solução:** política de *equal marginal utility*:

$$u'(E_t) = \lambda \quad \forall t.$$

*Intuição:* distribuir capacidade explicativa equitativamente entre decisões.

## AL.3. Métrica de Transparência

Define-se:

$T$  = tokens decisórios / tokens explicativos.

A EAAS impõe  $T_{\min} \geq 0.15$ : pelo menos 15% de uma resposta IA deve ser justificativa e não apenas preditiva.

---

# Epílogo Matemático e Filosófico do Whitebook EAAS

## 1. Síntese Técnica

A EAAS foi projetada, em seu núcleo conceitual, para:

- **Unificar** ERP, CRM, Marketplace, IA e Omnichat em um único ecossistema;
- **Aprender** continuamente por auto-reforço e regressão semântica;
- **Garantir** coerência, segurança e ética via restrições LTL + CMDP;
- **Balancear** custo, velocidade e precisão por duais  $(\lambda, \mu)$ ;
- **Isolar** domínios multi-tenant mantendo consistência de dados e conhecimento.

---

## 2. Síntese Matemática

O modelo completo combina:

- Espaços semânticos vetoriais  $(R^d)$  com grafos  $(G=(V,E))$ ;
- Funções Lipschitz e convexidade garantem estabilidade;
- Regularização e metadualidade asseguram equilíbrio ético e econômico;
- Regras de parada (Lyapunov) previnem loops cognitivos;
- Teoremas de percolação e calibração protegem robustez e precisão.

---

## 3. Síntese Filosófica

A IA da EAAS não busca substituir humanos, mas **compor inteligência coletiva**.  
Matematicamente, é uma função contínua sobre o espaço de intenções humanas —  
filosoficamente, é o *espelho lógico* do pensamento organizacional.

“Toda informação é uma semente: o que cresce dela depende do solo da mente que a recebe.”  
— *Fillipe Guerra, Epílogo do EAAS Whitebook*.

## 4. Conclusão Formal

### Teorema (Integração Cognitiva):

Se cada módulo  $M_i$  do EAAS é:

1. Estável (Lipschitz),
2. Ético (satisfaz LTL),
3. Explicável ( $T \geq 0.15$ ), e
4. Autoaprendente ( $\partial E / \partial t \geq 0$ ),

então o sistema global  $S = \cup_i M_i$  é **autônomo, auditável e coerente**.

*Prova conceitual:* segue da composição de funções estáveis sob constraints éticos e derivadas positivas de aprendizado.

---


## Agradecimento e Fecho

O *Whitebook EAAS – Everything As A Service*

é a expressão formal da convergência entre matemática, ética e tecnologia.


Ele demonstra que a verdadeira autonomia da IA não nasce do código, mas da **integração disciplinada entre raciocínio simbólico, inferência probabilística e responsabilidade humana**.

---

 **Autor Conceitual e Criador da EAAS:**  
Fillipe Guerra

 **Título completo:**  
**EAAS Whitebook — A Revolução Cognitiva Integrada da Plataforma Everything as a Service**

 **Ano:** 2025

 **Status:** Versão 1.0 – Final Conceitual