



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

## IA\_REVIEW\_PARTE3

1 mensagem

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

27 de outubro de 2025 às 18:21

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

Pra te ajudar a aplicar sem tropeços, deixo um **checklist de implantação rápido + smoke test**. Copia/cola e vai ticando:

### Checklist de implantação (IA v2 multimodal)

#### 1. Deps

- onnxruntime-node, hnswlib-node, multer, jsdom, pdf-parse, mammoth, iconv-lite, node-fetch, sharp, tesseract.js, unzipper, papaparse, dayjs, chart.js, react-chartjs-2.

#### 2. Modelos

- models/minilm.onnx, models/vocab.txt
- models/mobilenet.onnx (ou ViT), models/imagenet\_labels.json

#### 3. Migrations

- schema.ai.core.ts, schema.ai.graph.ts, schema.ai.eval.ts (tudo criado/aplicado).

#### 4. Config (.env)

- PRIMARY\_TENANT\_ID, AI\_\* (dims, paths, ANN, OCR, fresh half-life).

#### 5. IA v2

- Substituir pasta /server/ai/\* e /shared/schema.ai\* pelos arquivos da v2.
- Registrar rotas: registerAiRoutesV2, registerIngestExtendedRoutes, registerMetricsRoutes, registerKnowledgeRoutes.

#### 6. Páginas Admin

- /admin/metrics, /admin/entities, /admin/knowledge (Next.js + Tailwind).
- Registrar Chart.js no \_app.tsx (conforme lib).

#### 7. Pasta de dados

- Garantir data/ann/ com permissão de escrita (checkpoint HNSW).

#### 8. Build & run

- Subir servidor e verificar logs: inicialização ONNX + criação ANN.

### Smoke test (rápido, via curl)

# 1) Ingest de texto

```
curl -s -X POST http://localhost:3000/api/ai/ingest \
-H "Content-Type: application/json" \
-d '{"source":"manual","uri":"kb://intro","title":"Intro","text":"A Yes You Deserve é uma plataforma ...","meta":{"sourceRank":0.9}}'
```

# 2) Ingest de URL (HTML/PDF/DOCX/PPTX/CSV)

```
curl -s -X POST http://localhost:3000/api/ai/ingest.url \
-H "Content-Type: application/json" \
-d '{"url":"https://exemplo.com/algun.pdf}"'
```

# 3) Consulta (score completo  $\alpha/\beta/\gamma/\delta/\zeta$ )

```
curl -s -X POST http://localhost:3000/api/ai/ask.full \  
-H "Content-Type: application/json" \  
-d '{"query":"Quero um tour em Sintra com compra imediata", "k":12}' | jq '.[] | {score, parts, id}'
```

## Sanidade rápida

- **ANN:** após ingestir, verifique se criou/atualizou data/ann/<tenant>.\*.bin/.meta.json.
- **OCR:** suba uma imagem com texto em PT/EN → confira se criou doc source:"ocr".
- **Métricas:** endpoint /api/ai/eval/daily deve responder (após interações e job).
- **Painel:** acesse /admin/metrics e /admin/entities (top relações surgem conforme ingest).
- **Knowledge Explorer:** cole 2–3 URLs em /admin/knowledge, pré-visualize e aprove as melhores.