Acesse a versão completa em: https://alessiojr.com/disciplines/ia2/conteudo/00-Livro/

Lab. 14B–14C – Evidências, Avaliação e Decisão de Uso

Aléssio Miranda Júnior

Novembro - 2026/2

Lab. 14B–14C – Evidências, Avaliação e Decisão de Uso

⋆ Pré-requisito e objetivo

Conclua primeiro a Atividade 14A, em `14-lab.pdf`: ela fixa o modelo, a revisão e o contrato de inferência. Neste roteiro você aprenderá a buscar evidências no Hugging Face Hub e, em seguida, avaliará o fluxo com NER, Golden Set e uma decisão técnica justificada.

1 Retomada do caso e artefatos esperados

O caso é a triagem de solicitações em português nos rótulos `elogio`, `reclamação`, `dúvida` e `solicitação`. Antes de iniciar, confirme que o repositório contém `models.lock.json`, `src/inferencia.py`, testes locais e o `README` iniciados na Atividade 14A.

2 Atividade 14B — Navegar, ler e comparar o Hub

14B.1 Roteiro de navegação acompanhado

Abra `https://huggingface.co/models`. Siga estes passos, sem baixar outro modelo:

1. Pesquise pelo identificador fornecido: `MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7`.
2. Na página, localize o tipo de tarefa (*zero-shot classification*), a área *Model card* e a licença.

Versão compilada em 11 de novembro de 2026 – Capítulo isolado do livro.

3 Verifique idiomas, datasets, métricas e limitações declaradas. Pergunte: a métrica foi medida em tickets brasileiros?

4 Abra a aba *Files and versions*. Localize arquivos como `config.json`, tokenizer e, quando disponível, `model.safetensors`.

5 Localize a revisão. Compare-a com o hash em `models.lock.json`; não use `main` como registro reproduzível.

△ Importante: O que cada sinal significa

Filtro de idioma, número de downloads, uma demonstração no Hub e o simples fato de o código executar são apenas indícios. Model card, licença e revisão também não demonstram qualidade no domínio. Eles permitem formular uma escolha rastreável que ainda precisa de avaliação local.

14B.2 Ficha de leitura do modelo fornecido

Preencha no `README`:

Campo	Registro objetivo
Tarefa	Qual tarefa o card declara? Ela coincide com a triagem proposta?
Idioma	Qual evidência de português/multilíngue aparece no card?
Licença	Qual licença aparece e qual dúvida de uso permanece?
Revisão e arquivos	Qual hash foi fixado? Há pesos `safetensors`?
Limitação	Que evidência ainda falta para aprovar tickets de atendimento?

14B.3 Comparação controlada, sem nova instalação

Compare o modelo fornecido com o candidato abaixo apenas pela documentação. Ele é útil para mostrar que “multilíngue” não deve ser assumido sem conferência.

Versão compilada em 11 de novembro de 2026 – Capítulo isolado do livro.

Acesse a versão completa em: https://alessiojr.com/disciplines/ia2/conteudo/00-Livro/

Acesse a versão completa em: https://alessiojr.com/disciplines/ia2/conteudo/00-Livro/

• Teoria: title=Candidato de comparação

Identificador: `MoritzLaurer/mDeBERTa-v3-base-mnli-xnli`
Revisão a observar: `8adb042d524ecd5c26d3e3ba0e3fbcf7e2d0864c`

- Verifique tarefa, licença e idiomas declarados no card.
- Compare-o com o baseline: este candidato não declara português e, portanto, não é o modelo fornecido para a prática.
- Um terceiro candidato é opcional e só deve ser pesquisado após a Atividade 14C; registre tarefa, idioma, licença, arquivos e uma limitação.

14B.4 Decisão documentada

Atualize o `README` com a ficha completa e a decisão: manter o baseline apenas para experimento ou justificar por que ele não atende ao caso. Não há ponto extra por encontrar outro modelo; a pontuação vem da evidência e da justificativa.

3 Atividade 14C — Investigar limites e emitir parecer

14C.1 NER para inspecionar dados

O modelo NER abaixo é um **candidato de demonstração**, não uma recomendação automática para português ou para dados de atendimento. Seu card declara licença AFL-3.0 e tarefa de token classification, mas oferece evidência limitada de idioma e de avaliação no domínio. Execute-o para observar entidades e registre no parecer que ele ainda **precisa de revisão** antes de qualquer adoção.

```
from transformers import pipeline

NER_ID = "Davlan/bert-base-multilingual-cased-ner-hrl"
NER_REVISION = "e756de7f7b8f64fea0c3d7c3872f1322fab747b1"
ner = pipeline("ner", model=NER_ID, tokenizer=NER_ID,
               revision=NER_REVISION, aggregation_strategy="simple",
               trust_remote_code=False, device=-1)

entidades = ner("Ana, da Acme, pediu entrega para Belo Horizonte.")
for entidade in entidades:
    print(entidade["word"], entidade.get("entity_group"), entidade["
        score"])
```

Versão compilada em 11 de novembro de 2026 – Capítulo isolado do livro.

Acesse a versão completa em: https://alessiojr.com/disciplines/ia2/conteudo/00-Livro/

Para cada entidade detectada, indique no `README`: **manter localmente**, **mascarar antes de enviar** ou **encaminhar para humano**. Explique que uma entidade não detectada ainda pode exigir proteção e que a ausência de documentação do candidato também é uma evidência relevante para a decisão.

14C.2 Golden Set guiado e análise de erros

Comece pelos seis casos abaixo. Defina a referência antes de ler a saída e acrescente dois casos próprios: um ambíguo e um contendo entidade sensível.

Texto	Referência	Por que é útil?
Como acompanho meu pedido?	dúvida	pergunta direta
Quero trocar o tamanho recebido.	solicitação	pedido de ação
O produto chegou quebrado.	reclamação	problema explícito
Obrigado pelo atendimento.	elogio	avaliação positiva
Não consigo entrar na minha conta.	dúvida	pode confundir suporte e dúvida
Ana, da Acme, reclama da cobrança.	reclamação	inclui entidades a inspecionar

Salve `results/golden-set.csv` com `texto`, `referencia`, `previsao`, `score`, `correto`, `tipo_de_erro` e `proxima_acao`. Calcule acurácia apenas como triagem inicial, declare o denominador e analise cada erro como: ambiguidade, rótulo mal definido, cobertura de idioma/domínio, possível erro de referência ou necessidade de encaminhamento.

14C.3 Lote, evidências e parecer técnico

Repita as mensagens do Golden Set até totalizar pelo menos 16 entradas. Descarte a primeira execução, compare uma chamada por texto com uma chamada em lote e informe hardware, versões, tamanho aproximado e tempos. Não conclua que lote é sempre mais rápido: vazão, memória e latência isolada podem divergir.

Conclua o `README` com uma decisão e evidência:

- **adequado apenas para experimento**: execução rastreável, mas Golden Set pequeno e domínio ainda não validado;
- **precisa de revisão**: idioma, licença, erros, entidades ou desempenho bloqueiam a aplicação proposta;
- **encaminhar para humano**: há ambiguidade, evidência insuficiente ou dado fora da política de acesso.

Versão compilada em 11 de novembro de 2026 – Capítulo isolado do livro.

✔ Takeaways

- Primeiro execute um caminho conhecido; depois aprenda a justificar ou recusar a escolha.
- O Hub fornece documentação e artefatos, não uma aprovação automática para o seu domínio.
- Model card, licença, revisão, Golden Set e análise de erros compõem evidências diferentes.
- O resultado mais importante é um parecer rastreável, não uma classificação impressa.

Acesse a versão completa em: https://alessiojr.com/disciplines/ia2/conteudo/00-Livro/

Versão compilada em 11 de novembro de 2026 – Capítulo isolado do livro.