Fluxo de Fornecimento de Dados para Pesquisa

IA onde agrega valor, sempre com Humano no Loop

UNESP – CTInf

10 de novembro de 2025

1 Princípios

- **Finalidade e minimização**: entregar *apenas* o necessário ao objetivo aprovado.
- **Privacidade por desenho**: detecção de PII, desidentificação e checagem de egress como trilho padrão.
- **Humano no Loop (HITL)**: decisões de base legal, risco e liberação de saída são indelegáveis.
- **Rastreabilidade e reprodutibilidade**: *data cards*, *model cards* e lineage.

2 Fluxo original

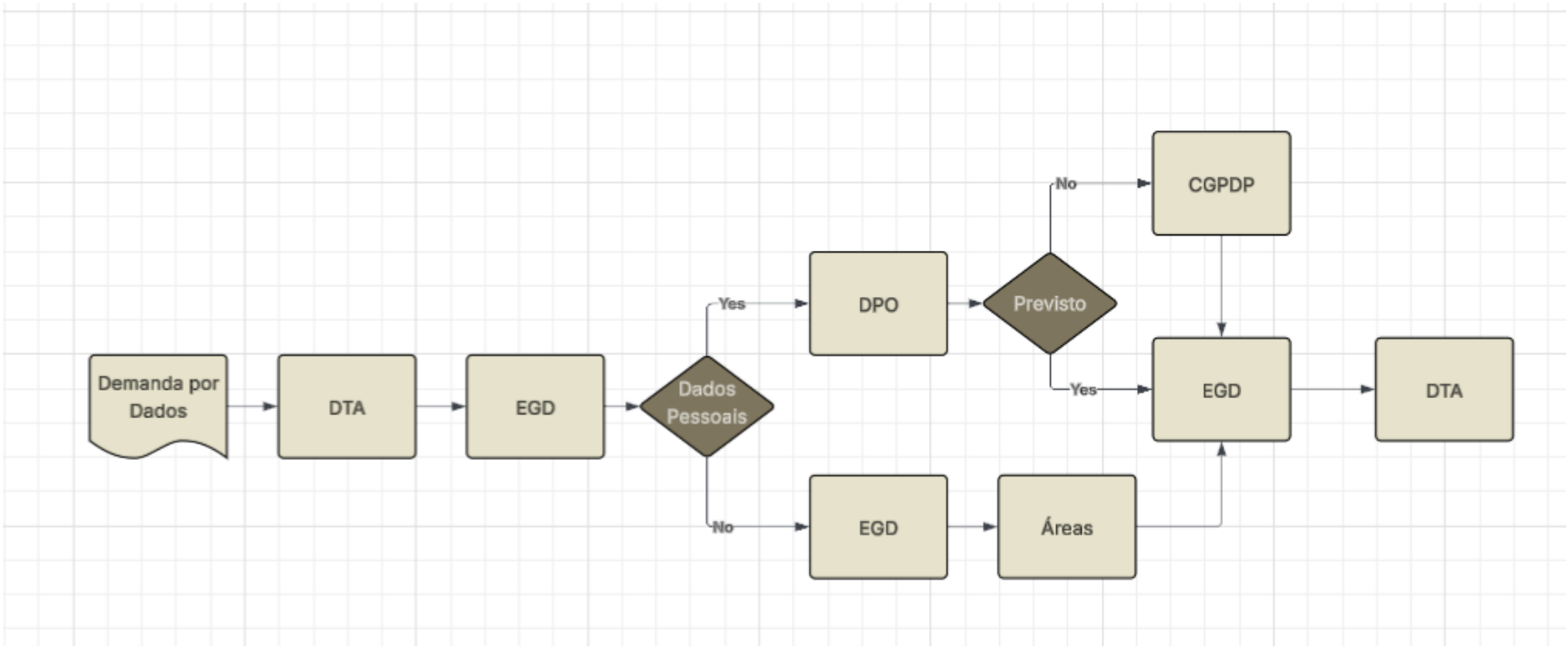


Figura 1: Fluxograma do processo original de fornecimento de dados para pesquisa.

2.1 Análise crítica do fluxo inicial (com DTA = Diretoria Técnica Acadêmica)

Leitura do diagrama. Partindo de *Demanda por Dados*, o pedido segue para a **DTA (Diretoria Técnica Acadêmica)** e, em seguida, para o **EGD** (governança). Há um *gateway* decisório **"Dados Pessoais?"**.

- **Se SIM:** o pedido vai ao **DPO** (encarregado de dados) e passa pelo nó **"Previsto?"**.
  - **Previsto = Não:** escala ao **CGPDP** e retorna ao **EGD**.
  - **Previsto = Sim:** retorna diretamente ao **EGD**.
- **Se NÃO:** o **EGD** aciona as **Áreas** detentoras dos dados, consolida e retorna ao **EGD**.

Em ambos os casos, o **EGD** conclui a preparação e devolve à **DTA**, que realiza o *fechamento/entrega*.

Forças do desenho.

a) **Roteamento por sensibilidade** ("Dados Pessoais?") reduz risco de tratamento indevido.

b) **Duplo nível de conformidade** para dados pessoais: *operacional* (DPO) e *estratégico* (CGPDP em exceções).

c) **EGD como *hub***: centraliza governança e reduz canais informais de acesso a dados.

d) **Ciclo com Áreas**: responsabiliza a fonte do dado e permite checagem de dicionários/definições.

Lacunas observadas.

a) **Critério "Previsto?"** pouco definido: sem referência explícita a registros de tratamento (ROPA), políticas de retenção e bases legais autorizadas.

b) **Ausência de trilhos técnicos** de privacidade/qualidade: não se explicitam anonimização/pseudonimização, critérios de agregação e verificação de vazamento de PII.

c) **Checkpoints humanos (HITL) não formalizados**: quem assina, quando e com qual evidência documental.

d) **Entrega orientada a arquivo**: não há menção a *Safe Room* (ambiente controlado) nem a *egress check* antes de saídas.

e) **Métricas e SLA** ausentes: não se medem prazos por etapa, taxa de devolução e motivos de negativa.

f) **Caminhos especiais** não representados: *CEP/CONEP*, transferências internacionais, dados de terceiros/convênio e casos de dados *mistos* (pessoais + sensíveis + não pessoais).

Recomendações pragmáticas (sem alterar a espinha dorsal).

1) **Definir objetivamente o nó "Previsto?"**: vincular a um catálogo/registro de tratamentos aprovado (ROPA) com base legal, finalidade, retenção e *pacotes de dados* predefinidos.

2) **Inserir trilho técnico mínimo entre EGD e Áreas/DPO**:
   - Classificação de sensibilidade (PII/PHI/confidencial/interno/público).
   - Política de tratamento: anonimização/pseudonimização/agregação com parâmetros (p.ex., k-anonymity, supressão, *top-coding*).
   - Validação de qualidade e vazamento de PII (regras e testes automatizados).

3) **Safe Room como padrão de entrega**: acesso em ambiente controlado (VDI/K8s), com *egress review* antes de qualquer exportação.

4) **Formalizar HITLs via RACI**: DTA (intake/SLA), EGD (governança/qualidade), DPO (base legal/risco), Áreas (fonte/dicionário), CGPDP (exceções) com artefatos obrigatórios (formulário, parecer, *data card*, termo de responsabilidade).

5) **Integrar CEP/CONEP** quando aplicável (intervenção com seres humanos ou dados sensíveis de saúde/biométricos).

6) **Implantar métricas de processo**: tempo médio por etapa, % "Previsto=Sim/Não", % devoluções por incompletude, incidentes de egress, backlog por tipo de dado.

Exemplos operacionais no nó "Previsto?".

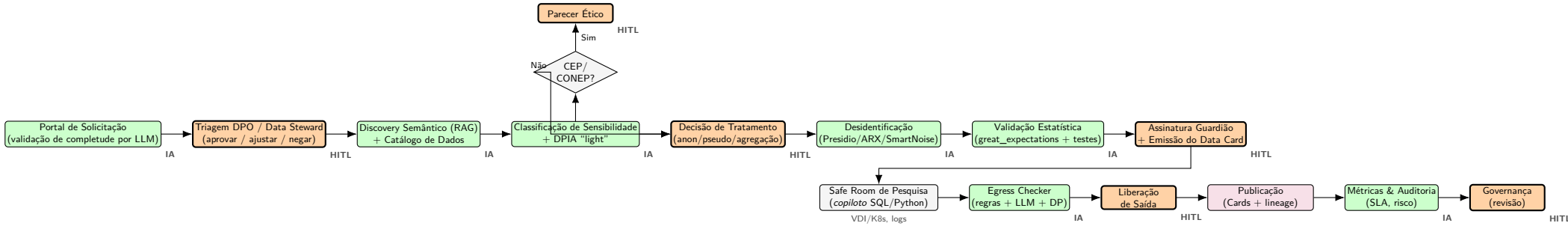
- **Previsto = Sim**: relatório semestral de evasão com coortes agregadas (política P–EDU–03; base legal de pesquisa – art. 7º/11º, §4º, LGPD), pacote padronizado no catálogo.
- **Previsto = Não**: vincular dados de saúde ocupacional a desempenho nominal individual para ranking; demanda *ad hoc* com CGPDP e, em regra, forte anonimização ou reprovação.

Efeito prático para a DTA (Diretoria Técnica Acadêmica).

- **Previsibilidade** do intake: formulários completos, rotas claras e SLAs definidos.
- **Velocidade sem risco**: com *Safe Room* e *egress check*, a DTA pode liberar acesso mais cedo e concentrar a discussão na saída (onde o risco é mensurável).
- **Rastreabilidade institucional**: artefatos (pareceres, *data cards*, logs) reduzem dependência de memória organizacional e facilitam auditorias.

Síntese. O fluxo atual já separa sensibilidade, centraliza no EGD e escala exceções ao CGPDP, o que é correto. Para fechar as brechas sem aumentar atrito, recomenda-se: (i) objetivar o nó *"Previsto?"* com catálogo/ROPA; (ii) instituir um trilho técnico mínimo de privacidade e qualidade; (iii) adotar *Safe Room* com *egress* obrigatório; e (iv) formalizar HITLs e SLAs. Assim, a DTA mantém velocidade operacional com segurança jurídica e científica.

3 Fluxo proposto (visão geral)


Parecer Ético
HITL
Sim
Não
CEP/
CONEP?
Portal de Solicitação
(validação de completude por LLM)
IA
Triagem DPO / Data Steward
(aprovar / ajustar / negar)
HITL
Discovery Semântico (RAG)
+ Catálogo de Dados
IA
Classificação de Sensibilidade
+ DPIA "light"
IA
Decisão de Tratamento
(anon/pseudo/agregação)
HITL
Desidentificação
(Presidio/ARX/SmartNoise)
IA
Validação Estatística
(great_expectations + testes)
IA
Assinatura Guardião
+ Emissão do Data Card
HITL
Safe Room de Pesquisa
(copiloto SQL/Python)
VDI/K8s, logs
Egress Checker
(regras + LLM + DP)
IA
Liberação
de Saída
HITL
Publicação
(Cards + lineage)
Métricas & Auditoria
(SLA, risco)
IA
Governança
(revisão)
HITL

4 Etapas e pontos com IA (e respectivos HITLs)

1. **Portal de Solicitação (IA + HITL-1)**: LLM privado checa completude e *data minimization*. DPO/Steward aprova/ajusta.
2. **Discovery Semântico (IA) + Validação (HITL-2)**: RAG no catálogo sugere datasets; Steward confirma adequação/restrições.
3. **Classificação de Sensibilidade & DPIA light (IA) + Decisão (HITL-3)**: PII/risco; decisão por anonimização, pseudonimização ou agregação.
4. **Desidentificação & Validação (IA) + Assinatura (HITL-4)**: Presidio/ARX/SmartNoise; testes de qualidade; guardião assina o *data card*.
5. **Safe Room com Copiloto (IA) + Auditoria (HITL-5)**: copiloto SQL/Python on-prem; auditoria amostral de consultas.
6. **Egress Checker (IA) + Liberação (HITL-6)**: regras + LLM para checar PII residual e *privacy budgets*; aprovação humana obrigatória.
7. **Publicação & Cards (IA) + Curadoria (HITL-7)**: rascunho automático de *data/model cards*; bibliotecário de dados revisa.
8. **Métricas & Auditoria (IA) + Governança (HITL-8)**: SLA, incidentes e risco monitorados; comitê ajusta políticas.

5 Ferramentas sugeridas

- **LLM privado**: Azure OpenAI ou Llama 3.1 afinado on-prem, com guardrails.
- **Catálogo & RAG**: BigQuery Data Catalog/Amundsen + Weaviate/PGVector.
- **Detecção/Anon**: Microsoft Presidio, ARX; *differential privacy* com SmartNoise.
- **Qualidade**: `great_expectations`.
- **Orquestração/Acesso**: Airflow/Prefect; Safe Room (VDI/K8s), API Gateway (Kong).
- **Políticas/Lineage**: OpenPolicyAgent; OpenLineage/Marquez.

6 Exemplos práticos (UNESP)

Evasão Acadêmica

PII (RA/nome) $\rightarrow$ pseudonimização + *k-anonymity*. Safe Room libera apenas agregados por coorte; egress bloqueia células com contagens < 15.

Saúde Ocupacional

Campos CID/atestados marcados como sensíveis $\rightarrow$ agregação + *differential privacy*. Parecer ético (CEP/CONEP) quando aplicável; publicação restrita.

7 Métricas de sucesso

- SLA por etapa; devoluções por incompletude; redução de *scope creep*.
- Risco de divulgação no egress; incidentes (meta: zero); satisfação do pesquisador.
- Aderência a *data cards/model cards*; cobertura de lineage.

8 Glossário (siglas e termos)

Sigla/Termo	Descrição
AI/IA	Inteligência Artificial; técnicas que aprendem padrões para auxiliar tarefas humanas.
API	Interface para comunicação entre sistemas.
ARX	Ferramenta open-source de anonimização (k-anonymity, l-diversity, t-closeness).
Azure OpenAI	Hospedagem corporativa de LLMs com controles.
BigQuery Data Catalog	Catálogo de metadados do Google.
CEP/CONEP	Comitê de Ética em Pesquisa / Comissão Nacional de Ética em Pesquisa.
CTInf	Coordenadoria de Tecnologia da Informação (UNESP).
DLP	Data Loss Prevention; detecção e proteção de PII/PHI.
DOI	Identificador persistente para dados/publicações.
DP	Differential Privacy (Privacidade Diferencial).
DPIA	Data Protection Impact Assessment (avaliação de impacto).
DPO	Encarregado pela proteção de dados (LGPD).
Egress Checker	Verificador de saídas do ambiente seguro (bloqueia PII residual).
Great Expectations	Testes/regras de qualidade de dados.
HITL	Human-In-The-Loop (checkpoint humano obrigatório).
K8s	Kubernetes; orquestração de contêineres.
Lineage	Rastreabilidade da origem/transformações/destino dos dados (OpenLineage/Marquez).
Kong	API Gateway para autenticação, auditoria e políticas.
LGPD	Lei Geral de Proteção de Dados (Brasil).
LLM	Large Language Model (copiloto/assistente).
Model Card	Ficha de propósito/dados/limites/risco de um modelo.
OPA	Open Policy Agent; autorização por políticas.
PHI	Protected Health Information (saúde).
PII	Personally Identifiable Information (dado pessoal identificável).
Presidio	Toolkit da Microsoft para detecção/mascaramento de PII.
Prefect/Airflow	Orquestradores de pipelines de dados/ETL.
RACI	Responsible, Accountable, Consulted, Informed (papéis).
RAG	Retrieval-Augmented Generation; busca + geração com LLM.
Safe Room	Ambiente seguro (VDI/K8s) sem saída de dados crus.
SLA	Acordo/indicador de nível de serviço.
SmartNoise	Biblioteca de privacidade diferencial.
spaCy	Biblioteca NLP para detecção de entidades/PII.
SQL	Linguagem padrão para consultas relacionais.
UNESP	Universidade Estadual Paulista.
VDI	Virtual Desktop Infrastructure (desktop virtual isolado).
Weaviate/PGVector	Armazenamento vetorial para catálogos/RAG.

Nota prática: os pontos HITL (1–8) no diagrama marcam decisões indelegáveis — base legal, aceitação de risco, liberação de egress e curadoria final. A IA acelera validações e documentação, mas **não** substitui a responsabilidade institucional.