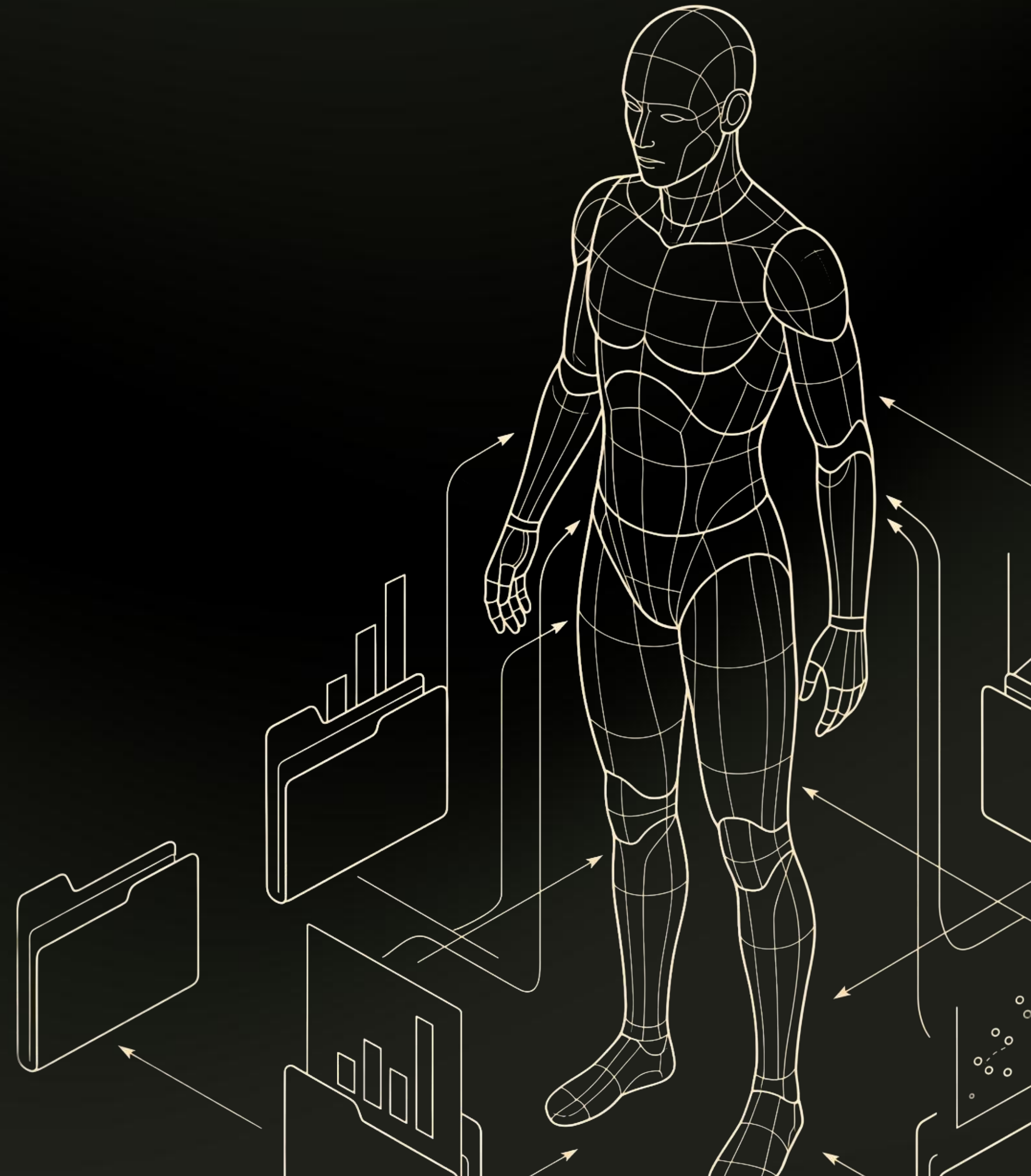




THUANY SERPA / SPECIALIST PRODUCT DESIGN

Usuários Sintéticos

Como otimizar seu discovery
validando hipóteses com AI_



Oi 🖐️

Eu sou a Thuany

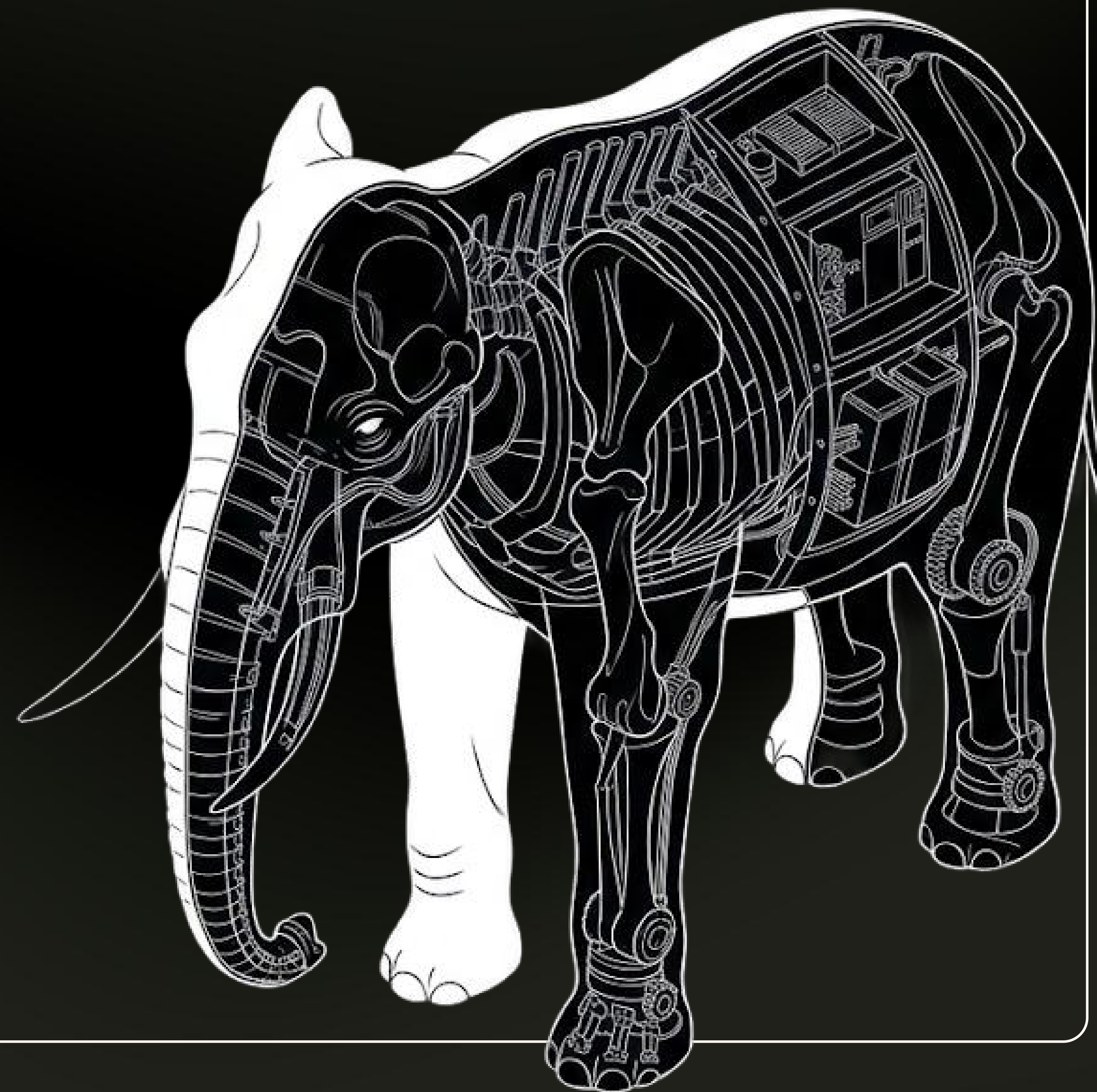
- ↳ +10 anos no design, com passagens por **SumUp**, **PicPay**, **Zé Delivery** e atualmente na **Cora**
- ↳ Pós graduada em Product Management e Inteligência Artificial
- ↳ Colecionadora de vinis e tutora de dois gatinhos



Vamos marcar um café?
tuserpa.me ↗



O elefante na sala das empresas AI-First





O problema do produto moderno

Hipóteses demais, ciclos de pesquisa de menos

ALGUMAS VERDADES DIFÍCEIS DE ENGOLIR

-  O custo do build costumava obrigar a hipótese a passar pela quali e pela quanti antes de virar produto.
-  Com a AI, o desenvolvimento ficou barato e esse freio sumiu. O teste pula direto pro produto como POC, e quem decide o que fica no ar é o próprio usuário.





Isso é user-centricity?

Usuário confuso, recebendo muitas atualizações, muitos produtos em beta e deixando de funcionar, deixando de utilizar recursos, etc

O aplicativo do [redacted] tá fora do ar a quase uma hora e eu so queria comprar [redacted]

mais alguém com esse problema na aba [redacted]? simplesmente não carrega nem aparece opções já entrei e sai diversas vezes da conta e não volta

Expectativa: uma **atualização** útil
Realidade: os donos desse hospício lançando o próximo pior update da história.
Como eles conseguem errar tanto no básico.
Custava fazer só o feijão com arroz?

Essa bomba desse **aplicativo** atualizou e agora ta tudo travando e dando erro. Toda nova mudança é pra piorar, impressionante

*Exemplos de reclamações em redes, relacionada a apps grandes, de empresas que estão escalando com AI.

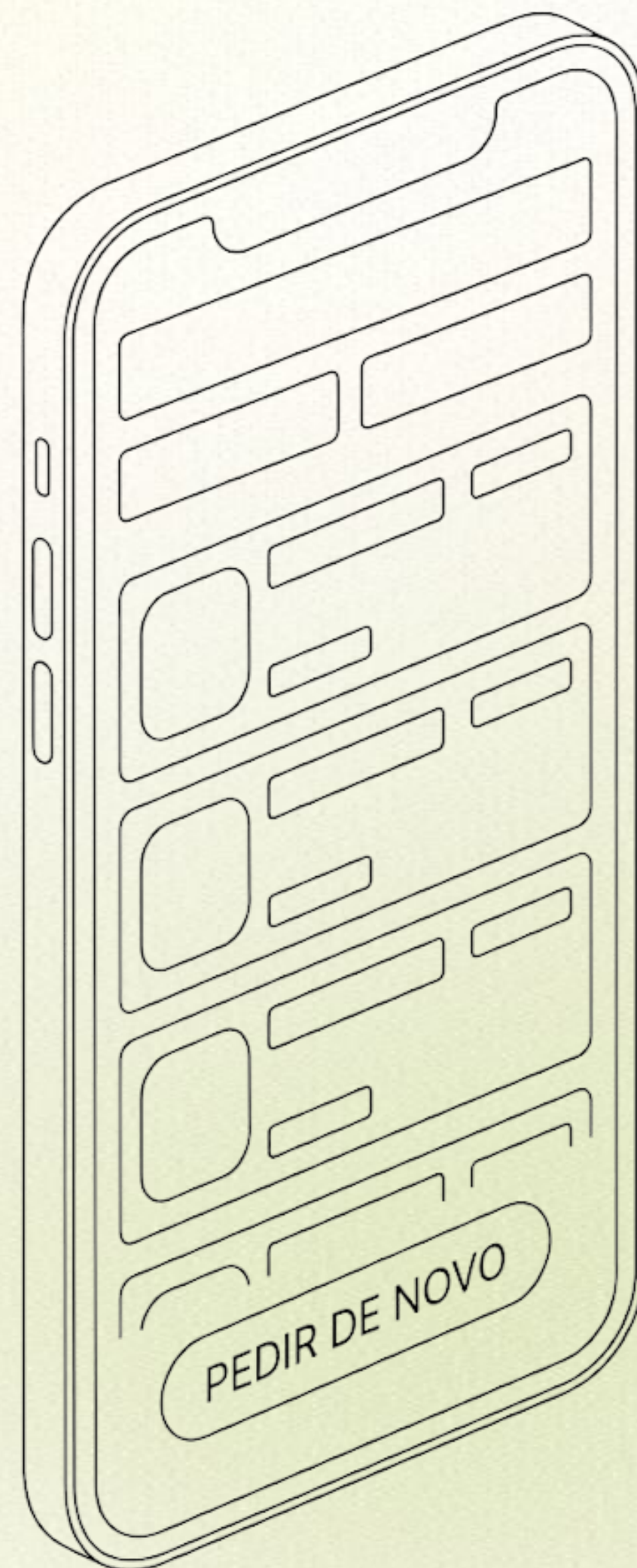


E se pudéssemos

... levar ideias mais refinadas
ao usuário sem travar a
experimentação?



A pergunta não é “validar mais rápido”. É recolocar um filtro que não dependa do custo do desenvolvimento pra existir.



CASE EXEMPLO

App de delivery de um restaurante_

“Se ‘Pedir de novo’ for o CTA primário na home, o cliente novo aumenta recompra em 7 dias.”



Esse fio acompanha a palestra inteira: vamos montar um modelo simples conversacional com dados reais para validar essa hipótese antes do MVP



Motivos para usar sintéticos **como filtro** (não substituto) de pesquisa

TEMPO



A hipótese chega ao ciclo de pesquisa já crivada por dado real – a entrevista aprofunda, não descobre do zero se a direção faz sentido.

Não gastar 2 semanas para descobrir que o CTA confunde quem ainda explora o cardápio

DINHEIRO



Recrutamento, incentivo e síntese ficam reservados pra perguntas que já sobreviveram a um primeiro crivo.

Recrutamento só se a hipótese sobreviver ao sintético

BASE DE CLIENTES



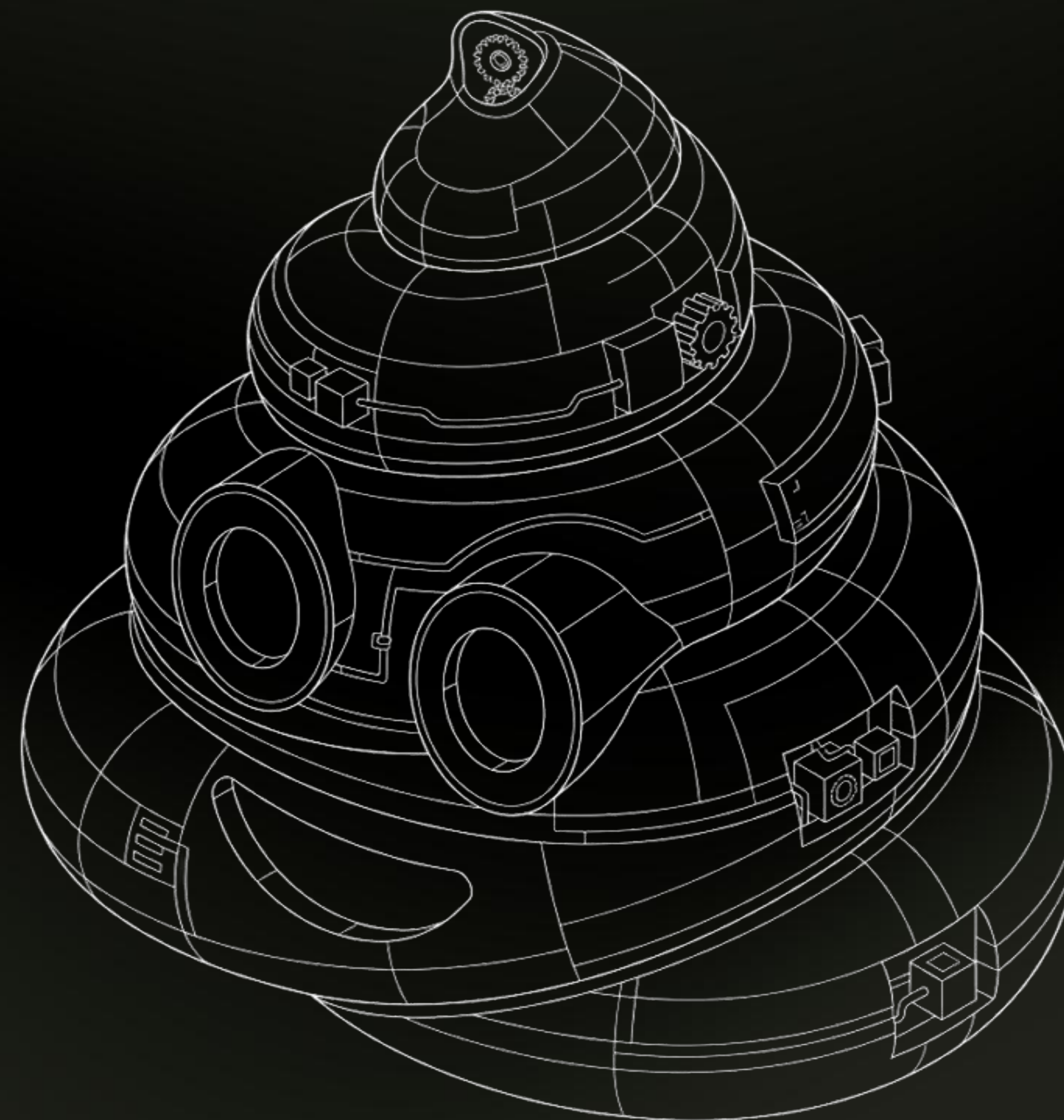
Sua base é finita e desgastável. O sintético evita gastar boa vontade validando o que poderia ter sido filtrado antes.

Não queimar cliente fiel testando atalho óbvio demais



Entra lixo, sai lixo

Vamos cuidar primeiro da camada de dados





A conversa só é **boa** quando temos as **evidências disponíveis**

Quatro camadas alimentam o índice do cluster. Sem elas, a síntese improvisaria.
Com elas, a resposta aponta pra fonte.

1 - QUEM

Demográfico / Firmográfico

Cadastro e contexto

ex.: 34 anos · zona sul · 2
pedidos · veio do Instagram

2 - NÚMEROS

Dados quantitativos

**frequência, ticket,
recência, conversão**

ex.: 5 sessões / 2 pedidos · ticket
R\$ 62 · 0 favoritos · 2 abandonos

3 - DECLARA

Dados qualitativos

**entrevista, reviews,
tickets de suporte**

ex.: “quero ver o cardápio sem
pressa” · “só descubro o frete no
final”

4 - FAZ

Comportamental

**eventos, sessões,
hesitações, abandonos**

ex.: scroll longo → compara itens
→ drop no frete → volta pra
home · Hotjar · Mixpanel · GA4



PREPARANDO OS DADOS

Embedding

É necessário que a informação qualitativa esteja buscável, trackeável e que tenha coerência. Isso é possível através do modelo embedding, que consiste na transformação dos textos em vetores.

1. TEXTO BRUTO

“quero ver o cardápio sem pressa”

2. VETOR

[0.82, -0.31, 0.15, 0.47, ...]

3. POSIÇÃO

um ponto no espaço, perto de quem
“quer dizer” algo parecido

exploração sem pressa

“pratos novos”

“sem pressa”

“olhar tudo antes”

“taxa escondida”

“frete surpresa”

frustração com o frete

eficiência e repetição

“já sei o que quero”

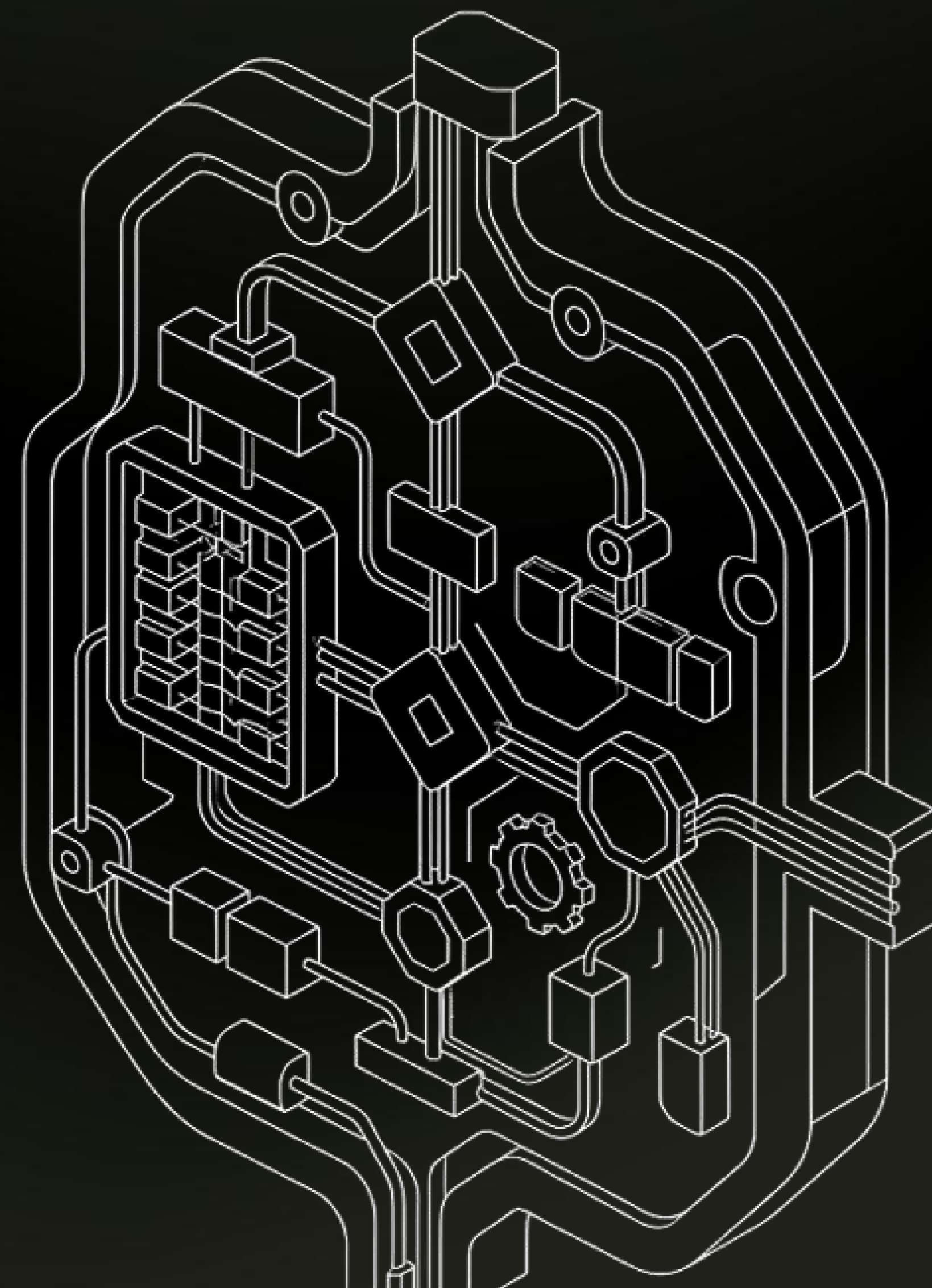
“quero atalhos”

Nenhuma dessas falas compartilha as mesmas palavras. O embedding aproxima quem tem a mesma intenção e afasta quem não tem. É isso que torna o dado qualitativo buscável e trackeável.



O mínimo de Machine Learning

que você precisa saber





Persona nasce de ~~suposição~~.

O sintético nasce de dados

QUADRO BRANCO

“Acho que existem três tipos”

Ana

32 anos
solteira,
prática

Bruno

40 anos
ocupado,
pai

Carla

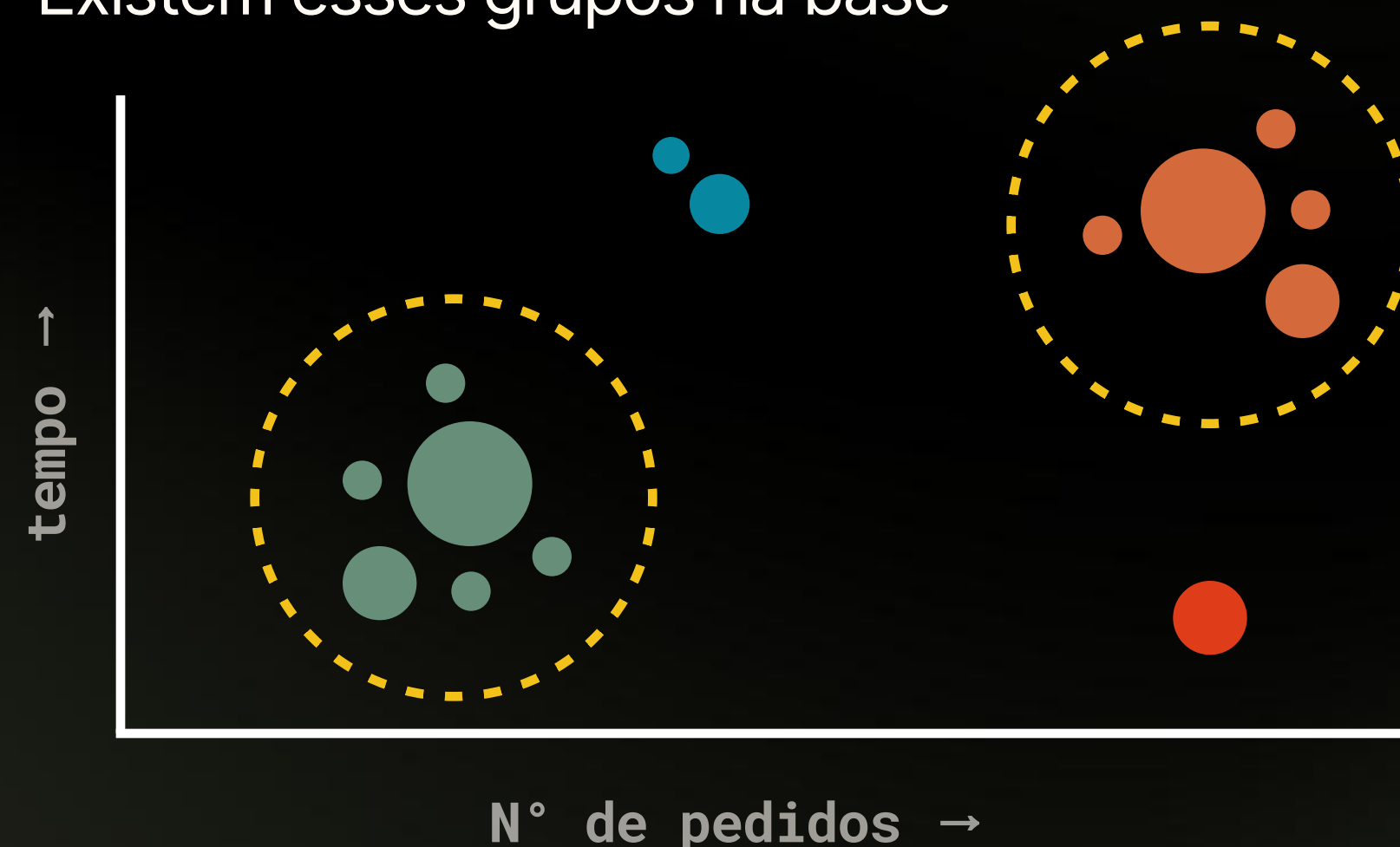
19 anos
estudante,
foodie

Fonte: amostra, vozes da cabeça, qualitativo

Personas projetadas - Forma imposta, imutável

CLUSTERIZAÇÃO

“Existem esses grupos na base”



Forma real que emerge - auditável e mutável



K-Prototype acha o grupo, DBSCAN acha quem foge dele



K-PROTOTYPE

Número + Categoria em uma única distancia

Atributo	C1	C2
Pedidos (num.)	<3	>12
Prato (cat.)	Varia sempre	Repete 1 - 2
Horário (cat.)	Aleatório	Fixo - Almoço

C1 → Explorador novo

menos de 3 pedidos, troca de categoria, sem favorito

C2 → Habitual do almoço

mais de 12 pedidos, 1 - 2 pratos fixos, sempre no almoço

DBSCAN

Densidade separa padrão de ruído



C3 → Outlier

ticket alto, frequencia alta, abandono crônico no checkout



Criando os sintéticos

Transformando o perfil e a voz
em um modelo virtual





1/4

Traduzindo um cluster em perfil

O cluster é um grupo de pessoas com padrões parecidos. Antes de conversar, o sistema resume o que mais repete sem fingir que todos são iguais

EXEMPLO - SEGMENTO EXPLORADOR

Padrão predominante: que define o segmento

> de 3 meses de uso · compara opções ·
ainda não criou favoritos · hesita quando
o custo aparece tarde_

Origem: cadastro + métricas + eventos.

O CUIDADO TÉCNICO

Predominância não significa universalidade

Parte do grupo pode agir diferente. O perfil guarda
variação, outliers e campos sem evidência.

Resultado: um arquétipo auditável, não uma pessoa
média inventada.



2/4

Transformando o perfil em um JSON

O JSON organiza o arquétipo num formato que o sistema consegue ler, versionar e usar em toda conversa.

EXEMPLO DE JSON

```
{
  "persona_id": "explorador_v1",
  "profile": {
    "lifecycle": "novo",
    "goal": "explorar com segurança"
  },
  "behavior": {
    "pattern": ["explora", "compara"],
    "frictions": ["custo tardio"]
  },
  "voice": {
    "register": "direto",
    "retrieval_scope": "explorador"
  },
  "boundaries": {
    "unknown": "baixa_confianca"
  }
}
```

Traduzindo o JSON

persona_id

nome do cluster de forma informativa

profile

quem é e o que busca

behavior

como costuma agir e onde trava

voice

qual conjunto de falas reais pode ser recuperado

boundaries

quando não há dado suficiente para responder

💬 O JSON não contém "a verdade do usuário". Contém instruções rastreáveis para a simulação.



3/4

Recuperando a voz com RAG^{*}

O perfil diz como o grupo age. O qualitativo mostra como ele fala sobre isso.

O RAG conecta os dois na hora da pergunta.

CHAT

Por que mostrar o preço por último gera atrito?

A pergunta vira embedding e busca apenas no qualitativo do segmento explorador.

FALAS RECUPERADAS

"Só descubro o custo quando já escolhi tudo."

"Prefiro comparar antes de me comprometer."

Fonte e data seguem junto com cada fragmento.



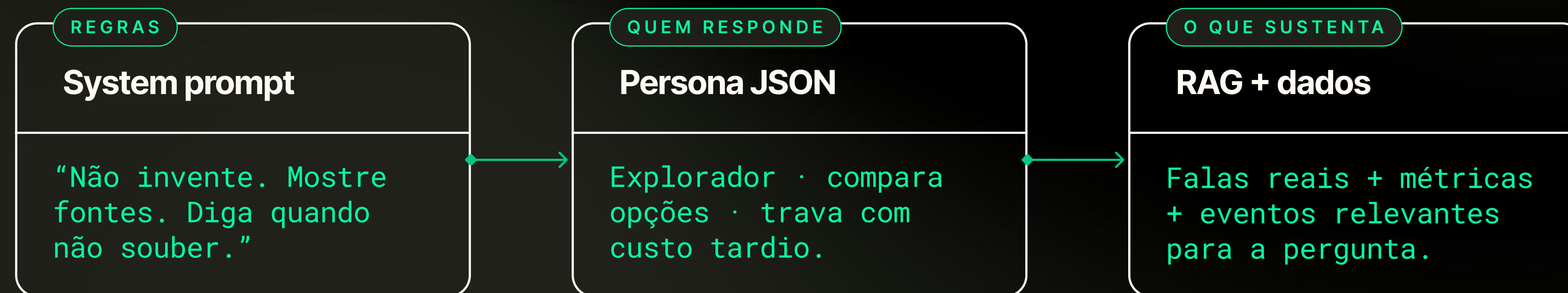
RAG (Retrieval Augmented Generation) é uma técnica utilizada para ampliar a capacidade de resposta de LLMs, combinando o conhecimento interno do modelo de linguagem com sistemas de recuperação de informações.



4/4

Montar o pacote que alimenta o LLM

O modelo não recebe só “responda como um explorador”. Ele recebe peças com funções diferentes.



_ pronto para receber perguntas

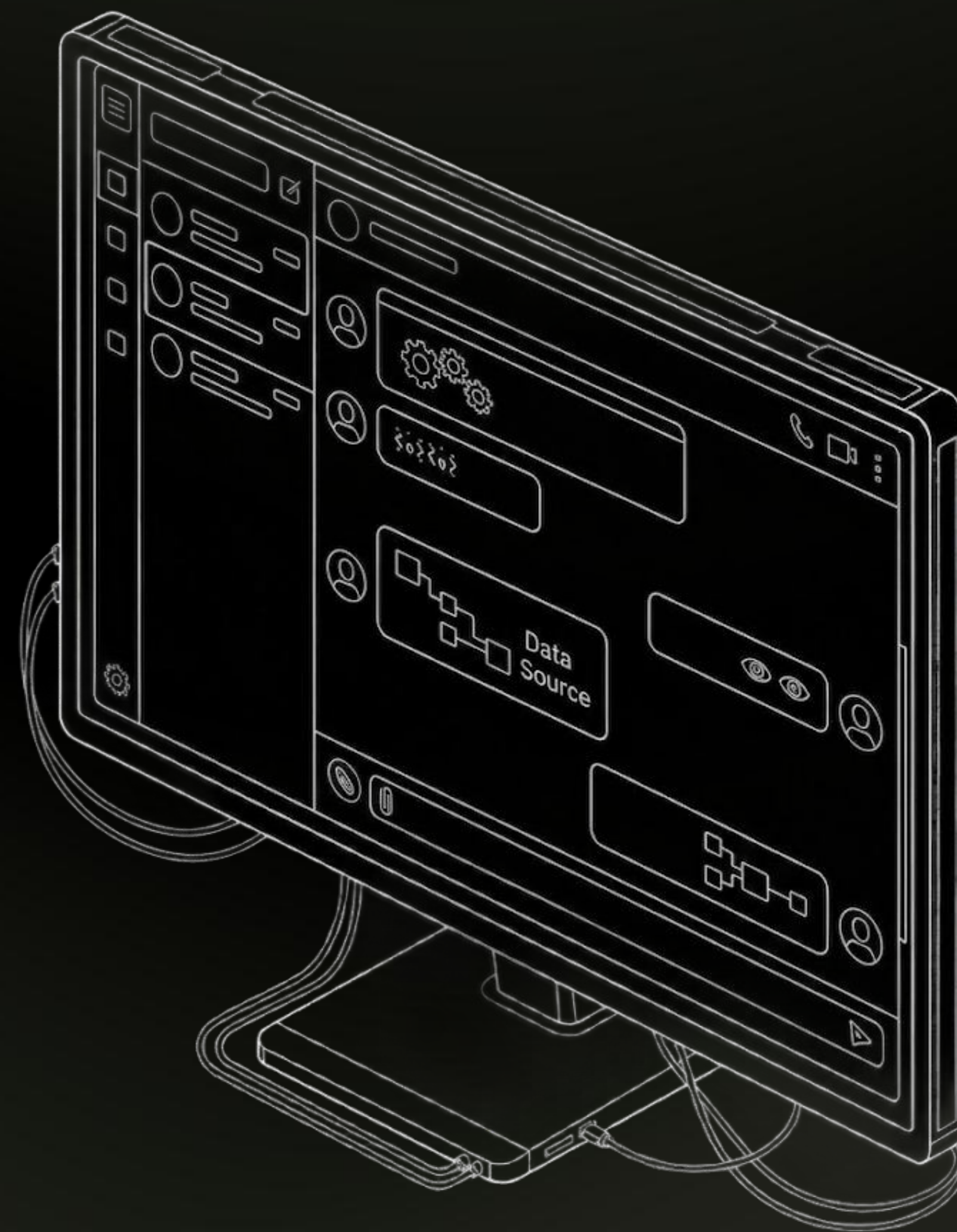
regras + JSON + evidências + pergunta → resposta simulada

O LLM interpreta o pacote. Não cria a persona nem sua história do zero.



Conversando com os dados

Usando os usuários sintéticos, e melhorando as funcionalidades e modelo





SCREENING

Antes de conversar, escolha o recorte

Nem toda pergunta cabe em toda a base. O screening calibra quem pode responder com assertividade – e quanto da amostra isso representa.

1. Pergunta

Hipótese em linguagem natural
“Pedir de novo” como CTA muda a decisão de compra?
2. Calibragem

Filtros do cluster
Pedidos, onboarding, sessões, região com população?
3. Saída

% assertividade
Escolhe os melhores clusters e abre o chat nesse recorte

US Synthetic
Discovery Lab

Buscar persona...

MAIN MENU

Screening

Chat

Síntese

REPORT

Cobertura

Personas

Case · Pedir de novo
Screening sintético para CTA da home

Screening

Calibre o recorte e escolha os melhores clusters

PERGUNTA DE PESQUISA

Se o botão “Pedir de novo” virar o CTA principal da home, isso mudaria sua decisão de compra?

Screening aponta o recorte da base com mais assertividade para essa pergunta.

3

Escolher os melhores clusters

CALIBRAGEM DO RECORTE

Qtd. de pedidos2+

Onboarding10 dias

Sessões/dia2/dia

Favoritos / repetição15%

PERÍODO DE VENDAS2

TERRITÓRIO · REGIÃO2

Resetar

Salvar cluster atual

USUÁRIOS6.291
recorte ativo

BASE10.000
amostra elegível

ASSERTIVIDADE DO RECORTE

63%

63% da amostra pode trazer assertividade nesta pergunta.

ConfiançaMÉDIA

CLUSTERS SALVOS

C1 · Explorador

78% · 7.800 · Explorador

2+10 dias2/dia15%

C2 · Habitual

71% · 7.100 · Habitual

12+21 dias3/dia65%

C3 · Sensível a custo

62% · 6.200 · Sensível a custo

6+14 dias2/dia40%

Abrir chat



CHAT

A pergunta para as personas

O sistema recupera

Perfil JSON · métricas do segmento · eventos · falas qualitativas sobre repetição, escolha e custo.

O sistema ainda não sabe

Reação real ao novo botão · efeito causal na recompra · emoção vivida diante da interface.

US

Synthetic
Discovery Lab

Q Buscar persona...

MAIN MENU

Screening

Chat

Síntese

REPORT

Cobertura

Personas

Case · Pedir de novo

Screening sintético para CTA da home

Chat

Converse com personas dos clusters salvos

Conversas · 6 personas

ALL

Todos

6 personas · 3 clusters

MC

Marina

Exploradora · 78% da base

RS

Rafael

Explorador · 78% da base

HR

Helena

Habitual · 71% da base

PM

Paulo

Habitual · 71% da base

DV

Diego

Sensível a custo · 62% da base

CN

Camila

Sensível a custo · 62% da base

Todos os clusters

Pergunta enviada a todos por padrão

← Screening

Síntese →

Marina · Exploradora

78%

Ainda comparo opções. Um CTA tão grande de 'Pedir de novo' assume que eu já escolhi.

Base: 78% · 7.800 usuários

Voz: comparar · confiança média

Rafael · Explorador

78%

Se ainda estou no cardápio, esse botão me empurra a repetir o que nem vi.

Base: 78% · 7.800 usuários

Voz: sem pressa · confiança média

Helena · Habitual

71%

Ajuda se mostrar o último pedido e o total com frete antes de confirmar.

Base: 71% · 7.100 usuários

Voz: o quê + quanto · confiança média-alta

Paulo · Habitual

71%

No almoço não tenho tempo. 'Pedir de novo' funciona se eu reconhecer o pedido.

Base: 71% · 7.100 usuários

Voz: atalho · confiança média

Para:

Todos

Marina

Rafael

Helena

Paulo

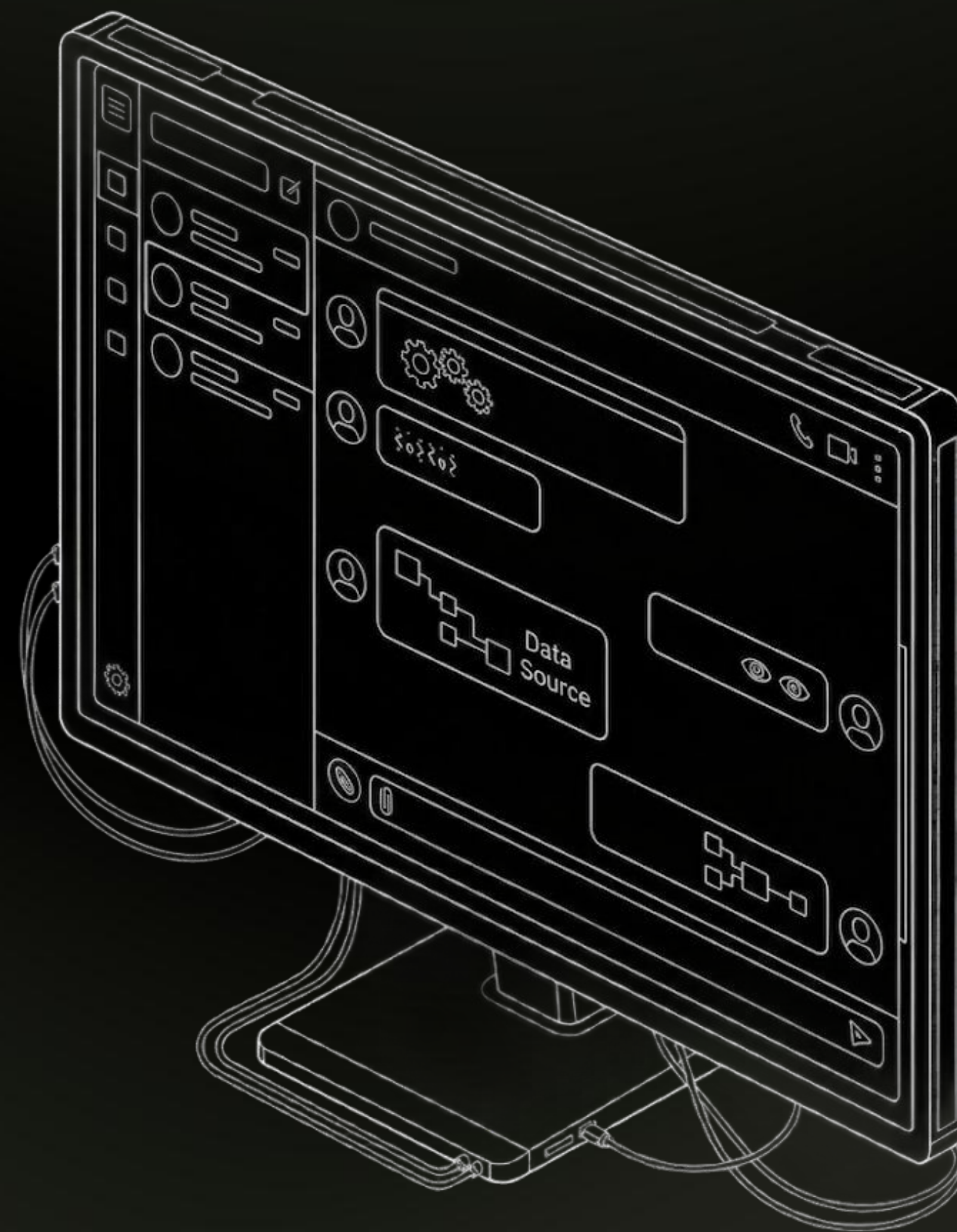
Escreva uma pergunta...

Enviar



Quando os dados apontam o sucesso

Qual resultado final é esperado após uma conversa com os dados





CHAT

Síntese da conversa

Com as personas calibradas e atualizadas é possível saber com mais certeza quando é momento de falar com o usuário real.

A síntese também pode ajudar a escolher quais são os públicos mais aderentes para campanhas de marketing, ou growth desde que tenhamos acesso aos dados

É possível integrar leitura de imagens, previsibilidade de comportamento do usuário em tela, análise heurística e de usabilidade, etc

US Synthetic Discovery Lab

Buscar persona...

MAIN MENU

Screening

Chat

Síntese

REPORT

Cobertura

Personas

Case · Pedir de novo

Screening sintético para CTA da home

Síntese

Resumo da conversa e decisão para a pesquisa

PERGUNTA DE PESQUISA

Se o botão “Pedir de novo” virar o CTA principal da home, isso mudaria sua decisão de compra?

Provavelmente atrita como CTA primário cego.

Direção de atrito com confiança média. Há condição de sucesso se o atalho mostrar total e pedido reconhecível.

RESUMO DA CONVERSA

C1 · Explorador

Ainda comparo opções. Um CTA tão grande de ‘Pedir de novo’ assume que eu já escolhi.

78%

7.800 usuários · confiança média-alta

C2 · Habitual

Ajuda se mostrar o último pedido e o total com frete antes de confirmar.

71%

7.100 usuários · confiança média-alta

C3 · Sensível a custo

Bloqueia sem frete e preço final visíveis.

62%

6.200 usuários · confiança média

VALE A PENA CONVERSAR COM USUÁRIO FINAL?

SIM — VALIDE COM GENTE

Sim. A síntese aponta direção, mas não mede reação à tela nem impacto em recompra. Entreviste/teste com o protótipo nos clusters de maior atrito e nos habituais condicionais.

SE SEGUIRMOS PARA POC / MVP

~39% → ~59%

CTA cego: ~39% de chance de a POC confirmar valor. Com total + frete + pedido reconhecível: sobe para ~59%. Ainda não é previsão causal — é probabilidade ilustrativa do screening.

Voltar ao chat

Screening



Sintético não substitui pesquisa. Torna a pesquisa real melhor e mais assertiva

TEMPO

hipótese interrogada
com personas antes do
campo

DINHEIRO

ciclo econômico com
teste pré validado no
ar, diminuindo ruído

BASE

usuário no centro
novamente, com
produto estável

O loop sempre volta pro humano. Agora ele volta no momento certo.



tuserpa.me

Obrigada!

