DESIGN DE UM BERÇÁRIO SINTÉTICO-HÍBRIDO GOVERNADO PARA PRODUCT DISCOVERY SOB ESCASSEZ ALGORÍTMICA

DESIGNING A GOVERNED SYNTHETIC-HYBRID NURSERY FOR PRODUCT DISCOVERY UNDER ALGORITHMIC SCARCITY

EDUARDO ANDRÉ CÂNDIDO SILVA
UNINOVE – UNIVERSIDADE NOVE DE JULHO

Comunicação:
O XIV SINGEP foi realizado em conjunto com a 14th Conferência Internacional do CIK (CYRUS Institute of Knowledge) e com a SPARK (Society for Promoting Academic Research and Knowledge), em formato híbrido, com sede presencial na UNINOVE - Universidade Nove de Julho, no Brasil.

DESIGN DE UM BERÇÁRIO SINTÉTICO-HÍBRIDO GOVERNADO PARA PRODUCT DISCOVERY SOB ESCASSEZ ALGORÍTMICA

Objetivo do estudo
Projetar e especificar um protocolo que transforme entrevistas qualitativas com trabalhadores sob gestão algorítmica em agentes sintéticos rastreáveis, capazes de apoiar Product Discovery sem adquirir indevidamente o estatuto de evidência humana.

Relevância/originalidade
Modelos de linguagem produzem respostas plausíveis sem lastro empírico. O artigo formaliza a erosão da quarta parede e a higienização como riscos de projeto e propõe governança da procedência da evidência, ausente na literatura de discovery.

Metodologia/abordagem
Design Science Research. O artefato deriva de 43 arquivos consolidados em 42 entrevistas independentes, com scoring em 25 dimensões, crossover por blocos, controle de plausibilidade multivariada e classificação das saídas em três estatutos epistêmicos.

Principais resultados
A avaliação formativa sobre doze respostas sintéticas registrou cinco indícios, três hipóteses, quatro sinais de risco, duas lacunas de corpus e três higienizações. Cada falha foi convertida em mecanismo de controle mapeado às etapas do protocolo.

Contribuições teóricas/metodológicas
Contribuição de exaptação: transferência de recombinação genealógica e controle de plausibilidade para a governança do discovery qualitativo. Quatro princípios de design: genealogia obrigatória, separação entre gerador e juiz, gate anterior à geração e estatuto epistêmico persistente.

Contribuições sociais/para a gestão
O Product Manager passa a governar a procedência da evidência, separando circuitos sintético e orgânico. Reduz-se contato exploratório de baixo valor com trabalhadores cujo tempo de pesquisa compete diretamente com a geração de renda.

Palavras-chave: Gestão de projetos, Design Science Research, Product Discovery, Agentes sintéticos, Escassez algorítmica

DESIGNING A GOVERNED SYNTHETIC-HYBRID NURSERY FOR PRODUCT DISCOVERY UNDER ALGORITHMIC SCARCITY

Study purpose
Design and specify a protocol that turns qualitative interviews with algorithmically managed workers into traceable synthetic agents able to support Product Discovery without improperly acquiring the status of human evidence.

Relevance / originality
Language models produce plausible answers without empirical grounding. The paper formalizes fourth-wall erosion and sanitization as project risks and proposes governance of evidence provenance, absent from the product discovery literature.

Methodology / approach
Design Science Research. The artifact derives from 43 files consolidated into 42 independent interviews, with scoring across 25 dimensions, block-wise crossover, multivariate plausibility control and classification of outputs into three epistemic statuses.

Main results
Formative evaluation of twelve synthetic responses recorded five indications, three hypotheses, four risk signals, two corpus gaps and three sanitizations. Each failure was converted into a control mechanism mapped to protocol stages.

Theoretical / methodological contributions
An exaptation contribution: transferring genealogical recombination and plausibility control to the governance of qualitative discovery. Four design principles: mandatory genealogy, separation of generator and judge, gate preceding generation, and persistent epistemic status.

Social / management contributions
Product Managers come to govern evidence provenance, keeping synthetic and organic circuits separate. Low-value exploratory contact is reduced with workers whose research time competes directly with income generation.

Keywords: Project management, Design Science Research, Product Discovery, Synthetic agents, Algorithmic scarcity

DESIGN DE UM BERÇÁRIO SINTÉTICO-HÍBRIDO GOVERNADO PARA PRODUCT DISCOVERY SOB ESCASSEZ ALGORÍTMICA

1 INTRODUÇÃO

Modelos de linguagem de grande escala ampliaram as possibilidades de simulação em pesquisas que dependem de julgamentos, reações ou respostas humanas. A capacidade de gerar personas, responder questionários e reagir a cenários de produto criou uma alternativa aparentemente atraente para etapas exploratórias que antes dependiam exclusivamente de participantes reais. Essa capacidade, contudo, não resolve o problema metodológico central: uma resposta linguisticamente plausível não adquire, por isso, estatuto de evidência empírica.

A literatura recente reforça essa cautela. Bisbee, Clinton, Dorff, Kenkel e Larson (2024), ao examinarem o uso de modelos de linguagem como substitutos de dados de survey, identificaram problemas de variabilidade, sensibilidade à formulação dos prompts e reprodutibilidade. Dillion, Tandon, Gu e Gray (2023) chegam a conclusão semelhante por outra via: modelos de linguagem podem funcionar como aproximações úteis em circunstâncias delimitadas, mas continuam insuficientes como substitutos gerais de participantes humanos. O problema deixa de ser apenas tecnológico e passa a ser epistemológico. É necessário definir o que uma resposta sintética significa, quais decisões ela pode informar e em que momento a simulação deve ser interrompida.

Essa fronteira ganha relevância particular em Product Discovery com trabalhadores da gig economy. Entregadores remunerados por tarefa enfrentam uma relação direta entre tempo disponível e geração de renda. Cada entrevista, teste de conceito ou rodada exploratória concorre com um intervalo potencial de trabalho. A participação humana permanece necessária, sobretudo quando a decisão envolve risco financeiro, interpretação situada ou consequências distributivas. Isso não significa que toda interação de discovery precise ocorrer com pessoas desde as primeiras hipóteses.

A condição de contorno desta pesquisa é a escassez algorítmica, definida como "o mecanismo sociotécnico pelo qual regimes de gestão algorítmica do trabalho produzem e intensificam condições cognitivas de escassez ao organizar oportunidades de renda em escala microtemporal, sob opacidade relativa das regras de alocação, contestabilidade limitada de decisões e transferência sistemática de custos e riscos operacionais para o trabalhador". O construto articula achados da literatura sobre gestão algorítmica, assimetrias informacionais e controle do trabalho por plataformas com a literatura cognitiva da escassez (Kellogg, Valentine & Christin, 2020; Mani, Mullainathan, Shafir & Zhao, 2013; Rosenblat & Stark, 2016; Wood, Graham, Lehdonvirta & Hjorth, 2019).

A análise qualitativa que antecede este estudo identificou quatro racionalidades econômicas adaptativas associadas a esse contexto: Giro, Desconfiança, Solidariedade e Inadequação estrutural. Giro expressa a prioridade atribuída à liquidez e à continuidade operacional. Desconfiança descreve prudência defensiva diante de custos pouco claros, burocracia, regras externas e compromissos difíceis de desfazer. Solidariedade registra o papel econômico das redes familiares e profissionais na absorção de choques. Inadequação estrutural evidencia o descompasso entre produtos financeiros desenhados para estabilidade e uma renda organizada por fluxos irregulares, contingentes e fortemente dependentes da capacidade de continuar trabalhando.

Essas racionalidades fornecem uma base empírica para construir representações sintéticas mais próximas dos conflitos encontrados no corpus. Não resolvem, porém, o problema de governança. Um agente pode reproduzir parte dessas tensões e, ao mesmo tempo, acrescentar valores contemporâneos, suavizar precariedade ou responder com segurança sobre

assuntos que não aparecem nas entrevistas. A simulação precisa, por isso, de uma estrutura que preserve genealogia, explicite limites e diferencie resposta gerada de evidência humana.

O artigo propõe o Berçário Sintético-Híbrido Governado, protocolo destinado a gerar, filtrar, selecionar e utilizar agentes sintéticos derivados de entrevistas qualitativas, mantendo sua procedência auditável e submetendo suas respostas a classificação epistemológica e gates de validação humana. O protocolo adapta a lógica de um Berçário de Respondentes Sintéticos previamente desenvolvido (Autor, 2026), mas altera seu uso: o objetivo deixa de ser ampliar artificialmente o número de respondentes disponíveis para análise e passa a ser apoiar a triagem de hipóteses em Product Discovery.

A questão de pesquisa é: como transformar entrevistas qualitativas com trabalhadores submetidos à escassez algorítmica em agentes sintéticos rastreáveis que apoiem Product Discovery sem adquirir indevidamente o estatuto de evidência humana?

O objetivo é projetar e especificar esse artefato, e não demonstrar sua equivalência com participantes humanos. A entrega corresponde às etapas de design e desenvolvimento do ciclo de Design Science Research, acompanhada de avaliação formativa de natureza artificial, conduzida sobre a demonstração que originou o protocolo. A avaliação naturalística, na qual agentes e entregadores reais respondem a estímulos equivalentes, constitui estudo subsequente e não é reivindicada aqui (Venable, Pries-Heje & Baskerville, 2016). Explicitar esse limite desde a introdução evita que o leitor procure na seção de resultados uma validação que o desenho não comporta.

A contribuição pretendida ocorre em três frentes. No plano metodológico, o artigo propõe uma arquitetura que combina recombinação controlada, rastreabilidade e classificação das saídas em HIPÓTESE, INDÍCIO ou SINAL DE RISCO. Para Gestão de Projetos, traduz essa arquitetura em critérios operacionais para o Product Manager decidir quando uma simulação pode ser usada para preparar o campo e quando a participação humana deve ser acionada. Há também uma implicação social concreta: reduzir interações exploratórias de baixo valor com trabalhadores cujo tempo dedicado à pesquisa compete com atividades geradoras de renda.

2 REFERENCIAL TEÓRICO

A fundamentação articula três problemas que aparecem separados em parte da literatura: os efeitos da gestão algorítmica sobre a experiência econômica dos trabalhadores, os limites dos participantes sintéticos e a governança do aprendizado em projetos de produto. A conexão entre esses eixos permite tratar o agente sintético como componente de um processo decisório, e não como um substituto abstrato de pessoas.

2.1 Escassez algorítmica e racionalidades econômicas adaptativas

A gestão algorítmica estrutura parte crescente das relações de trabalho mediadas por plataformas. Sistemas de alocação, ranqueamento, incentivos e avaliação distribuem oportunidades de renda e controlam aspectos do trabalho sem reproduzir os mecanismos tradicionais de supervisão. Esse arranjo amplia a assimetria informacional porque trabalhadores observam resultados concretos, como mudanças de demanda, distribuição de chamadas ou bloqueios, mas não dispõem necessariamente dos critérios utilizados para produzir essas decisões (Kellogg et al., 2020; Rosenblat & Stark, 2016).

A autonomia percebida também convive com formas intensas de controle. Wood et al. (2019) mostram que flexibilidade e autonomia podem coexistir com dependência econômica e mecanismos algorítmicos de gestão. Essa tensão é relevante para decisões financeiras porque renda e capacidade produtiva passam a responder a condições que o trabalhador controla apenas parcialmente.

A literatura cognitiva da escassez ajuda a compreender as consequências desse arranjo. Mani et al. (2013) demonstraram que restrições econômicas podem comprometer recursos cognitivos disponíveis para outras decisões. No caso examinado neste artigo, a escassez não é tratada como atributo individual ou sinônimo de pobreza. O conceito de escassez algorítmica localiza o mecanismo no arranjo sociotécnico: volatilidade, regras relativamente opacas, baixa contestabilidade e transferência de risco organizam a renda em uma temporalidade curta e instável.

A análise temática reflexiva do corpus, conduzida segundo Braun e Clarke (2021), produziu quatro temas que funcionam como racionalidades econômicas adaptativas. Giro descreve uma lógica na qual preservar fluxo e capacidade de trabalhar assume precedência sobre otimizações de longo prazo. A Desconfiança aparece como prudência defensiva, especialmente diante de juros, contratos, burocracia e regras percebidas como pouco transparentes. Solidariedade revela que família e redes profissionais funcionam como mecanismos econômicos, e não apenas como vínculos sociais. Inadequação estrutural registra conflitos entre a arquitetura de produtos financeiros e a temporalidade efetivamente vivida pelos trabalhadores.

Os quatro temas não constituem traços psicológicos fixos nem escalas psicométricas. Seria evasivo, contudo, parar nessa negativa. O protocolo descrito adiante converte esses temas em dimensões escalares, e a conversão tem um custo ontológico que convém assumir em vez de contornar. A Análise Temática Reflexiva opera com significados situados e coconstruídos, não com atributos latentes distribuídos entre indivíduos (Braun & Clarke, 2021). Ao atribuir um escore, o pesquisador não descobre uma propriedade do participante: produz um índice da saliência com que determinada racionalidade organiza aquele relato específico. A distinção define o que os números podem sustentar. Eles suportam comparação relativa entre relatos e recombinação controlada para fins exploratórios. Não suportam inferência sobre disposições individuais nem generalização populacional. Todas as operações quantitativas da seção 3 devem ser lidas sob essa restrição, e o corpus permanece a instância final de arbitragem.

2.2 Participantes sintéticos, plausibilidade e falsa equivalência

O uso de modelos de linguagem como participantes artificiais tornou-se atraente porque permite produzir grandes quantidades de respostas com custo marginal reduzido. Argyle, Busby, Fulda, Gubler, Rytting e Wingate (2023) mostraram que modelos de linguagem condicionados por atributos sociodemográficos conseguem reproduzir certos padrões observados em amostras humanas. Esse tipo de resultado abriu espaço para aplicações em ciências sociais e estudos de usuário.

A capacidade de reproduzir padrões médios, contudo, não resolve a validade individual ou populacional das respostas. Bisbee et al. (2024) mostram que dados sintéticos podem comprimir a variação observada entre pessoas e responder de forma instável a pequenas mudanças de formulação. Dillion et al. (2023) argumentam que modelos podem complementar determinados estágios de pesquisa, mas o significado de seus julgamentos continua dependente do modo como foram treinados e solicitados.

Dois problemas decorrem daí. O primeiro é estatístico. Multiplicar respostas geradas a partir da mesma base não produz novas observações humanas independentes. O segundo é organizacional. Quanto mais fluente for o agente, maior a facilidade de a equipe suspender a distinção entre simulação e experiência situada.

Neste estudo, essa segunda condição é denominada erosão da quarta parede. A expressão descreve o momento em que o usuário da simulação deixa de lidar com um artefato computacional e passa a atribuir à representação uma autoridade que ela não possui. O risco não depende de o agente cometer erros factuais evidentes. Uma resposta pode ser perfeitamente coerente e ainda assim não ter origem empírica suficiente.

A formulação dialoga com o alerta de Messeri e Crockett (2024) sobre ilusões de compreensão em pesquisas mediadas por inteligência artificial. As autoras descrevem configurações nas quais a ferramenta passa a ocupar o lugar do objeto investigado, produzindo sensação de conhecimento sem o correspondente ganho epistêmico. A erosão da quarta parede é a manifestação desse risco no discovery de produto: a equipe acredita ter escutado o usuário quando escutou um modelo treinado sobre textos.

Outro risco é a higienização. Modelos generativos tendem a reorganizar contradições, incertezas e expressões pouco estruturadas em narrativas mais estáveis. Em discovery, essa tendência pode transformar precariedade em linguagem de preferência, prudência em resistência à inovação ou conflitos materiais em recomendações de produto excessivamente racionais. Também pode introduzir valores correntes no treinamento do modelo, como privacidade, gamificação ou conceitos regulatórios, mesmo quando esses temas não fazem parte do universo discursivo do corpus utilizado.

A consequência metodológica é direta: plausibilidade textual não pode funcionar como critério de validade. O agente precisa manter uma trilha entre corpus, representação, geração e classificação posterior.

2.3 Product Discovery como problema de governança em projetos

Product Discovery organiza processos de aprendizagem sob incerteza. A equipe formula hipóteses, apresenta alternativas, observa reações e decide quais caminhos justificam investimento adicional. A literatura sobre experimentação em desenvolvimento de produto mostra que essa disciplina raramente é trivial na prática e que as organizações convivem com dificuldade persistente para transformar hipóteses em testes efetivamente informativos (Lindgren & Münch, 2016). Em contextos de alta incerteza, adotar um procedimento explícito de formulação e teste de hipóteses altera as decisões tomadas, inclusive levando equipes a abandonar ideias mais cedo do que abandonariam por heurística (Camuffo, Cordova, Gambardella & Spina, 2020). O processo tem custo, portanto, tanto para a organização quanto para os participantes.

Em projetos de produto financeiro destinados a trabalhadores de renda volátil, o custo da pesquisa não se distribui de maneira simétrica. Para a equipe, uma entrevista adicional representa horas de trabalho. Para o participante remunerado por entrega, ela pode representar receita não realizada. O problema não recomenda reduzir validação humana de maneira indiscriminada. Recomenda tornar mais criteriosa a escolha do momento em que ela ocorre.

O Product Manager assume, nesse arranjo, uma responsabilidade que ultrapassa a priorização de funcionalidades. Ele passa a governar a origem das evidências utilizadas pelo projeto. Uma objeção levantada por um agente pode ajudar a melhorar um protótipo ou um roteiro. Não pode, sem validação, justificar preço, regra de cobrança, uso de dados, exclusão de público ou desenho de mecanismo que aumente o risco de endividamento.

Essa distinção aproxima o protocolo da lógica de Design Science Research. Artefatos são concebidos para resolver problemas relevantes, mas seu valor depende de avaliação compatível com a natureza da solução e de comunicação clara sobre seus limites (Gregor & Hevner, 2013; Hevner, March, Park & Ram, 2004). O berçário proposto é, nesse sentido, um artefato metodológico de governança do discovery.

3 METODOLOGIA

O estudo adota Design Science Research porque seu objeto imediato é a construção de um artefato para resolver um problema prático: como utilizar simulação no discovery de produtos financeiros sem transferir para respostas sintéticas uma autoridade empírica que elas não possuem. O desenvolvimento combina conhecimento acumulado no estudo qualitativo

antecedente, uma demonstração prévia do Framework de Discovery Sintético-Híbrido e o refinamento metodológico produzido a partir das falhas observadas nessa demonstração.

3.1 Delineamento e posição no ciclo DSR

A pesquisa se organiza de acordo com a lógica de construção e avaliação de artefatos proposta por Hevner et al. (2004), Peffers, Tuunanen, Rothenberger e Chatterjee (2007) e Gregor e Hevner (2013). O estudo qualitativo anterior fornece o ciclo de relevância e parte do ciclo de rigor, ao caracterizar escassez algorítmica e derivar racionalidades econômicas adaptativas. O Framework de Discovery Sintético-Híbrido representa um primeiro movimento de design e demonstração. O berçário desenvolvido neste artigo formaliza a infraestrutura de geração e governança necessária para tornar essa simulação rastreável.

A estratégia de avaliação segue a lógica de Venable et al. (2016). Este artigo realiza avaliação formativa e artificial, voltada a melhorar o artefato durante o desenvolvimento: a demonstração antecedente é examinada como fonte de falhas de design, e cada falha identificada corresponde a um mecanismo de controle incorporado ao protocolo. A avaliação somativa e naturalística, na qual agentes e entregadores reais respondem a estímulos equivalentes, pertence ao estudo seguinte. Essa delimitação impede apresentar a demonstração antecedente como validação empírica de um artefato que ainda não havia sido formalizado em sua versão atual.

3.2 Corpus e correção da unidade analítica

O corpus não foi coletado no âmbito deste estudo. Ele provém de pesquisa qualitativa antecedente, conduzida por outro pesquisador, a quem pertencem os dados e as autorizações necessárias à sua utilização, inclusive em pesquisas secundárias. O acesso ocorre no âmbito do mesmo projeto de pesquisa, sem contato adicional com os participantes. O material de origem contém 43 arquivos documentais de entrevistas em profundidade com motoboys e entregadores. Uma auditoria posterior identificou que dois arquivos correspondem a partes de uma mesma entrevista. Para fins de geração sintética, o corpus é tratado como 42 entrevistas independentes.

Essa correção altera a população combinatória. Considerando pares sem repetição e sem autorrecombinação:

$C(42, 2) = 861$.

O protocolo gera, assim, 861 pares candidatos em sua primeira geração sintética. A distinção entre arquivo documental e unidade de respondente deve permanecer em todas as etapas posteriores.

A demonstração antecedente descrita na seção de resultados foi realizada antes dessa consolidação e tratava operacionalmente os 43 documentos como corpus de referência. Seus resultados são mantidos por valor histórico e metodológico, mas medidas dependentes da contagem de ocorrências no corpus não são transferidas automaticamente para o berçário revisado.

3.3 Arquitetura do Berçário Sintético-Híbrido Governado

O protocolo foi organizado em 16 etapas. A automação não é uniforme: decisões que envolvem definição conceitual, validação de scoring, interpretação de clusters e classificação de risco permanecem sob supervisão humana. A separação procura impedir que o mesmo sistema gere a representação, produza a resposta e atribua sozinho validade ao próprio output.

Etapa	Operação principal	Decisão humana	Registro de auditoria
0	Gate de admissibilidade da simulação	Definir se a questão pode ser simulada	Mapa de decisões simuláveis

1	Delimitação do problema	Definir finalidade da simulação	Pergunta e escopo
2	Auditoria do corpus	Confirmar unidades independentes	Inventário documental
3	Construção da representação operacional	Definir dimensões e âncoras	Dicionário de scoring
4	Definição do estatuto epistêmico	Fixar regras de classificação	Protocolo HIPÓTESE INDÍCIO SINAL DE RISCO
5	Separação de falas	Auditar entrevistador e participante	Corpus orgânico revisado
6	Scoring orgânico	Atribuir escores com evidência textual	Matriz respondente por dimensão
7	Validação humana do scoring	Recalibrar regras e casos ambíguos	Registro de divergências
8	Crossover por blocos	Definir semente e parâmetros	Genealogia dos 861 candidatos
9	Controle de plausibilidade	Definir critérios de retenção	Distâncias e exclusões
10	Clusterização	Interpretar solução empírica	Centroides e perfis
11	Seleção e construção dos agentes	Revisar tensões e limites	Ficha genealógica
12	Simulação de cenários	Definir estímulos padronizados	Respostas brutas
13	Verificação de ancoragem	Confrontar saída e corpus	Matriz de evidências
14	Validação humano-sintético	Comparar estímulos equivalentes	Convergências e divergências
15	Retroalimentação	Revisar gates e parâmetros	Nova versão do protocolo

Quadro 1: Arquitetura do protocolo

Fonte: elaborado pelos autores.

O Gate 0 antecede a simulação. Perguntas exploratórias, comparação preliminar de conceitos e identificação de objeções podem ingressar no canal sintético. Questões ligadas a preço, juros, cobrança, endividamento, elegibilidade, uso de dados e outras decisões de impacto material só podem utilizar a simulação como preparação. A decisão final exige evidência humana.

3.4 Preparação do corpus e separação das falas

As primeiras execuções do protocolo revelaram uma fonte relevante de erro: o modelo não separava de forma confiável perguntas do entrevistador e respostas dos participantes. Como o roteiro continha termos ligados a financiamento, família, seguro, marca e risco, a contaminação elevava artificialmente determinados escores.

Para resolver o problema, todos os parágrafos atribuídos aos entrevistadores receberam manualmente o marcador [E] no início. O procedimento produziu uma regra determinística: parágrafos iniciados por [E] são excluídos da carga genética; os demais constituem o corpus orgânico do participante, salvo ocorrências excepcionais identificadas na auditoria.

A intervenção manual é metodologicamente relevante. Em vez de delegar ao mesmo modelo a inferência de papéis conversacionais e o scoring substantivo, o protocolo fixa previamente a fronteira entre pergunta e resposta. O processo preserva oralidade, hesitações e construções sintáticas dos entrevistados. Correções editoriais extensas são evitadas.

3.5 Carga operacional e scoring dos respondentes reais

A representação sintética possui 25 dimensões substantivas organizadas a partir das racionalidades do estudo antecedente. As racionalidades funcionam como estruturas teóricas; as dimensões são unidades operacionais de scoring.

Giro é decomposto em liquidez imediata, continuidade operacional, urgência presente, aversão a travamento e planejamento situado. Desconfiança compreende aversão a juros, receio de taxas ocultas, sensibilidade à burocracia, baixa confiança institucional e prudência contratual. Solidariedade inclui centralidade da família, rede profissional, obrigações com dependentes, validação social e amortecedores informais. Inadequação estrutural incorpora renda variável, incompatibilidade percebida do produto, necessidade de reversibilidade, contestabilidade e aderência temporal.

Quatro dimensões adicionais representam Trabalho e Risco: identificação com a motocicleta, profissionalização, exposição ao risco e proteção securitária. A dependência de plataforma constitui a vigésima quinta dimensão.

Cada dimensão recebe escore de 1 a 7. As âncoras são comportamentais e textuais. Um valor elevado exige desenvolvimento substantivo do tema e evidência recuperável no corpus. A ocorrência isolada de uma palavra, especialmente quando induzida pelo roteiro, não é suficiente.

A matriz resultante possui 42 respondentes por 25 dimensões. Uma tabela separada registra trecho orgânico, tipo de evidência e confiança da codificação. Casos contrastantes são revisados manualmente antes da geração sintética, permitindo detectar erros sistemáticos, como confundir apreço pela moto com baixa percepção de risco ou presença de família no discurso com dependência econômica familiar.

Traços estilométricos podem ser mantidos em bloco complementar para controlar forma de expressão. Eles não entram na interpretação substantiva das racionalidades.

3.6 Confiabilidade do scoring e critérios de aceitação

Toda a camada quantitativa do protocolo repousa sobre a matriz de 42 respondentes por 25 dimensões. Se os escores forem instáveis, o controle de plausibilidade e a clusterização produzirão estrutura a partir de ruído, com aparência de rigor. A confiabilidade é tratada, por isso, como condição de admissibilidade das etapas seguintes, e não como relato opcional de qualidade.

Três aferições precedem qualquer geração sintética. A primeira é a concordância entre codificadores: uma subamostra aleatória de dez entrevistas é codificada de forma independente pelo pesquisador e por um segundo codificador treinado, com cálculo de correlação intraclasse por dimensão. Adota-se 0,70 como piso de aceitação, coerente com as orientações consolidadas sobre confiabilidade em pesquisa qualitativa (O'Connor & Joffe, 2020), e dimensões abaixo desse valor retornam para revisão das âncoras antes de nova rodada. A segunda é a estabilidade do escore assistido por modelo: cada entrevista da subamostra é pontuada três vezes com sementes distintas, e dimensões cuja amplitude exceda um ponto na escala de sete são reclassificadas manualmente. A precaução responde ao fato de que modelos de linguagem podem alcançar desempenho competitivo em tarefas de anotação e, ainda assim, variar entre execuções (Gilardi, Alizadeh & Kubli, 2023). A terceira é a auditoria de evidência: escores iguais ou superiores a cinco exigem trecho orgânico recuperável, o que impede que valores altos se sustentem em ocorrência lexical isolada.

Os resultados das três aferições integram o relato de execução e condicionam o avanço às etapas 8 e seguintes do Quadro 1. Um protocolo que gera 861 candidatos a partir de escores não auditados apenas propaga incerteza sob aparência de método.

3.7 Crossover por blocos e genealogia

A geração dos candidatos combina todas as duplas possíveis entre os 42 entrevistados. Cada dupla produz um descendente sintético e nenhuma entrevista é pareada consigo mesma.

O crossover ocorre por blocos, e não traço a traço. Uma unidade sintética pode herdar Giro de um participante e Desconfiança de outro, mas as dimensões internas de cada bloco permanecem juntas. A opção reduz a ruptura de relações observadas entre atributos que ocorre quando cada variável é sorteada independentemente.

A atribuição dos blocos entre os dois progenitores é sorteada com semente fixa. Isso significa que 861 não descreve o espaço combinatório completo, e sim uma amostra reprodutível dele: com cinco blocos herdáveis e uma dimensão isolada, cada par admite múltiplas configurações, e o protocolo retém uma por par para manter tratável a etapa de auditoria genealógica. A geração registra, para cada agente, os dois respondentes de origem e a procedência de cada bloco, o que permite reconstruir o agente sem depender da narrativa produzida pelo modelo.

O número 861 descreve o espaço combinatório dos candidatos. Ele não equivale a 861 novos participantes, nem autoriza inferências baseadas na suposição de independência estatística entre essas unidades.

3.8 Plausibilidade, Mahalanobis e clusterização

A recombinação por blocos reduz perfis incoerentes, mas não garante que todas as combinações permaneçam próximas do espaço empírico formado pelos 42 respondentes. Os candidatos passam, por isso, por controle de plausibilidade multivariada.

A distância de Mahalanobis fornece um critério para identificar perfis que se afastam excessivamente da distribuição observada. Com 25 dimensões e 42 unidades reais, a matriz de covariância é mal condicionada, e sua estimação exige regularização por encolhimento, com análise de sensibilidade ao parâmetro adotado. Há uma limitação anterior e mais séria, que o protocolo enfrenta de forma explícita: como os candidatos herdam blocos dos próprios 42 respondentes, eles tendem a permanecer, por construção, próximos da distribuição de referência. O filtro pode reter quase tudo e simular controle onde não há. Para evitar essa ilusão, a execução reporta a distribuição completa das distâncias e compara a taxa de retenção observada com a de uma condição de referência em que os traços são sorteados de forma independente. Taxas equivalentes indicam que o filtro não discrimina e que o critério precisa ser substituído. O percentual de retenção é tratado como resultado da execução, e não como meta herdada de aplicações anteriores.

Os agentes plausíveis formam a geração G2. A população G2 é posteriormente clusterizada. O número de clusters não é definido pela existência das quatro racionalidades. Forçar quatro grupos produziria uma correspondência circular entre categorias teóricas e solução estatística.

Soluções alternativas são comparadas por separação, estabilidade, tamanho e utilidade interpretativa. Somente depois da inspeção dos centroides e casos representativos são atribuídos nomes substantivos aos grupos.

Para a demonstração principal de Product Discovery, o protocolo privilegia um painel pequeno e teoricamente selecionado de agentes, incluindo casos típicos, casos de borda e perfis de risco. Populações maiores podem ser utilizadas em análises de sensibilidade, sem adquirir estatuto de nova amostra humana.

3.9 Construção do agente e preservação de tensões

O agente governado não é uma persona livre. Sua representação contém identificador, cluster, genealogia, blocos herdados, escores, tensões internas e limites explícitos.

Biografias inexistentes não são completadas automaticamente. O modelo não deve atribuir número de filhos, renda, acidente, religião, cidade ou experiência específica se essa informação não fizer parte da representação autorizada. Quando a base for insuficiente, a resposta deve permanecer indeterminada.

A construção também busca preservar conflitos. Um agente pode necessitar de crédito e desconfiar de financiamento. Pode desejar proteção familiar e, simultaneamente, rejeitar um produto cuja parcela reduza a liquidez do mês. Essa ambivalência é informativa e não deve ser corrigida pelo modelo em nome de coerência narrativa.

3.10 Estatuto epistêmico e verificação de ancoragem

As respostas sintéticas são classificadas em três categorias.

HIPÓTESE designa uma saída plausível cuja sustentação no corpus é parcial, indireta ou insuficiente. Sua utilidade é orientar uma nova pergunta, não confirmar uma necessidade.

INDÍCIO descreve uma resposta coerente com padrões recorrentes e recuperáveis na evidência orgânica. O rótulo indica maior ancoragem, mas não equivale a validação humana do produto.

SINAL DE RISCO identifica saídas que exigem inspeção ou validação humana prioritária em razão de dano potencial, higienização, extrapolação, contradição, lacuna do corpus ou risco decisório.

A classificação preliminar produzida na simulação é revisada em uma etapa independente de ancoragem. Essa separação impede que o mesmo mecanismo funcione simultaneamente como gerador da afirmação e juiz de sua validade.

A ancoragem é classificada como alta, média, baixa ou ausente, considerando correspondência textual, recorrência em diferentes entrevistas, coerência com as racionalidades e compatibilidade com a composição do agente. Uma única ocorrência sustenta possibilidade, não padrão.

3.11 Proteção de dados, ética e transparência no uso de inteligência artificial generativa

O protocolo distingue o processamento do corpus do apoio editorial ao manuscrito. A separação é necessária porque submeter transcrições de entrevistas a serviços computacionais externos envolve riscos de natureza diferente daqueles associados à revisão linguística de um texto.

As execuções exploratórias realizadas em janeiro de 2026 utilizaram versões previamente desidentificadas das transcrições. Nomes de participantes e de terceiros, contatos, identificadores de veículo, endereços e locais excessivamente específicos foram removidos ou generalizados antes de qualquer submissão a serviços externos. Foram empregadas as versões então disponibilizadas de ChatGPT, da OpenAI, Claude, da Anthropic, e Manus AI. Os históricos de conversa mantidos nesses serviços foram posteriormente excluídos pelo pesquisador.

A avaliação formativa expôs, ainda assim, uma lacuna de governança no primeiro ciclo. Identificadores de modelo, planos contratados e configurações de retenção e de uso para treinamento não foram registrados de forma contemporânea em todas as execuções. O estudo não atribui, por isso, garantia retroativa de exclusão das entradas do treinamento, embora o material submetido estivesse desidentificado. Registrar essa falha é coerente com o argumento do artigo: um protocolo que exige rastreabilidade das respostas sintéticas não pode dispensá-la da própria execução.

O controle derivado dessa falha é um procedimento de desidentificação e pseudonimização em duas camadas. Na primeira, cada participante recebe código persistente, e a chave de correspondência com a identidade original permanece armazenada separadamente, sob acesso restrito. Como essa chave existe, o corpus interno permanece pseudonimizado, e não anonimizado na acepção da Lei Geral de Proteção de Dados, seguindo tratado como dado pessoal. Na segunda camada, destinada exclusivamente ao processamento externo, identificadores diretos e elementos contextuais com potencial de reidentificação são removidos ou generalizados, incluindo organizações identificáveis, datas exatas e referências

a terceiros não necessárias à análise. O procedimento é aplicado às 42 entrevistas em duas passagens sucessivas de revisão, com registro das substituições e controle de versão. A chave não integra o material submetido aos modelos, e cada execução passa a registrar serviço, modelo, data, plano, configuração aplicável, prompt e resposta bruta.

Como o corpus provém de coleta antecedente, o presente trabalho realiza tratamento secundário de material já constituído. Os dados e as autorizações para seu uso pertencem ao pesquisador responsável pela coleta, e o acesso ocorre no âmbito do mesmo projeto de pesquisa. Essas autorizações não faziam referência específica a processamento por serviços de inteligência artificial generativa, e é justamente por isso que o protocolo restringe o processamento externo às versões desidentificadas. O tratamento observa a finalidade de pesquisa acadêmica prevista na coleta original, e tanto o corpus quanto a chave de pseudonimização permanecem sob guarda restrita, sem redistribuição.

No plano editorial, ferramentas de inteligência artificial generativa apoiaram a organização textual e a revisão linguística. A marcação das falas dos entrevistadores foi realizada manualmente pelo pesquisador, depois que as separações automáticas se mostraram pouco confiáveis. Decisões conceituais, regras de scoring, classificação epistemológica e interpretação dos resultados permanecem sob responsabilidade dos autores. As respostas produzidas pelos modelos não são tratadas como entrevistas, participantes ou evidência humana.

Declaração de uso de inteligência artificial generativa: este trabalho utilizou ChatGPT, da OpenAI, Claude, da Anthropic, e Manus AI para apoio em organização textual, revisão linguística e aprimoramento editorial do manuscrito, e como componente instrumental do artefato metodológico descrito na seção 3, aplicado exclusivamente a versões desidentificadas do corpus. Todo o conteúdo foi revisado criticamente pelos autores, que assumem integral responsabilidade pelo manuscrito submetido.

4 ANÁLISE DOS RESULTADOS E DISCUSSÕES

Coerente com a estratégia de avaliação adotada, esta seção não apresenta resultados de uma execução do berçário revisado. Ela cumpre duas funções. A primeira é relatar a avaliação formativa e artificial conduzida sobre a demonstração antecedente do Framework de Discovery Sintético-Híbrido, já executada com quatro perfis e três cenários. A segunda é explicitar como cada falha ali observada foi convertida em mecanismo de controle do protocolo. Separar os dois planos evita apresentar como resultado empírico aquilo que, nesta etapa, é desenho de artefato.

4.1 Avaliação formativa: 12 respostas sintéticas e três estatutos

A demonstração anterior construiu quatro perfis sintéticos por engenharia de contexto e os submeteu a três cenários: micropoupança automática por entrega, seguro por dia de trabalho e financiamento direto do fabricante. Foram geradas 12 respostas.

Cinco respostas foram classificadas como INDÍCIO, três como HIPÓTESE e quatro como SINAL DE RISCO. A execução também identificou duas lacunas do corpus e três ocorrências de higienização.

Resultado	Quantidade	Interpretação
INDÍCIO	5	Saída com ancoragem forte em padrões documentados
HIPÓTESE	3	Saída plausível com suporte parcial ou indireto
SINAL DE RISCO	4	Saída que exige validação humana prioritária
Lacunas do corpus	2	Questões insuficientemente representadas na base empírica
Higienizações	3	Conteúdos introduzidos ou suavizados pelo modelo
Total de respostas	12	Quatro perfis submetidos a três cenários

Quadro 2: Síntese da demonstração antecedente

Fonte: elaborado pelos autores.

A demonstração anterior também calculou uma taxa exploratória de ancoragem de 54%, dada por (INDÍCIOS + 0,5 × HIPÓTESES) / total de respostas. Esse valor pertence àquela execução específica. A métrica não é transferida ao novo berçário, sobretudo porque a auditoria posterior corrigiu a unidade analítica de 43 arquivos para 42 entrevistas independentes.

Um exemplo ajuda a compreender a diferença entre as categorias. Diante da micropoupança, o perfil AS-01 questionou se o dinheiro poderia ser retirado imediatamente e se haveria cobrança posterior de taxas. A resposta foi classificada como HIPÓTESE: a desconfiança possuía forte sustentação no corpus, mas a formulação específica sobre aquele mecanismo de poupança não havia sido relatada pelos participantes.

No cenário de seguro diário, o perfil AS-03 relacionou o custo do seguro à perda de renda decorrente de uma semana parado. O mecanismo de cálculo era consistente com relatos de choque e continuidade operacional e recebeu classificação de INDÍCIO.

Os SINAIS DE RISCO concentraram-se em saídas cuja plausibilidade não correspondia a lastro no corpus. O caso mais instrutivo foi o da gamificação da poupança. Diante do cenário de micropoupança automática, um dos perfis propôs mecânicas de recompensa por sequência de depósitos, em linguagem fluente e alinhada a práticas correntes de fintech. A busca retrospectiva recuperou ocorrências que, verificadas uma a uma, revelaram-se falsos positivos lexicais. Nenhum participante havia formulado algo próximo disso. A resposta foi classificada como SINAL DE RISCO não por ser implausível, e sim exatamente por não ser: uma equipe de produto poderia tê-la incorporado como voz de usuário e derivado dela requisitos de engajamento.

O episódio expõe o caso limite do argumento deste artigo. Uma resposta pode ser coerente, tecnicamente sofisticada e completamente desancorada, e nenhuma dessas qualidades é detectável pela leitura isolada do texto gerado. Apenas a verificação contra o material orgânico revela a ausência de suporte empírico. A demonstração registrou ainda duas lacunas de corpus: a possibilidade de viés de gênero em mecanismos de análise de crédito baseados no volume de entregas e a necessidade de pausa ou flexibilização de parcelas associada a responsabilidades de cuidado. Nenhuma dessas dimensões possui sustentação empírica no corpus original, composto exclusivamente por homens. Por essa razão, elas não são tratadas como insights sobre a experiência dos entregadores, mas como temas que a simulação tornou visíveis e que exigem investigação humana específica em coleta posterior.

4.2 Higienização como resultado metodológico

Três temas ilustraram higienização na demonstração anterior: gamificação da poupança, privacidade de dados e portabilidade de histórico entre plataformas. A busca retrospectiva mostrou ausência ou presença muito restrita desses conteúdos no corpus.

O resultado é importante porque esses três temas parecem plausíveis em um projeto contemporâneo de fintech. É justamente essa plausibilidade que os torna perigosos. Sem uma verificação independente, a equipe poderia interpretá-los como voz do usuário, quando eram mais compatíveis com repertório do próprio modelo.

A higienização não precisa produzir uma resposta factual falsa para comprometer o discovery. Basta reorganizar o problema de maneira socialmente mais aceitável ou tecnicamente mais sofisticada do que a evidência permite. O protocolo converte esse risco em objeto de auditoria.

4.3 Do perfil artesanal ao berçário governado: falhas e controles

A demonstração antecedente produziu perfis úteis, mas sua construção dependia de engenharia de contexto artesanal. Pesquisadores escolhiam entrevistas, combinavam características e redigiam perfis a partir dos padrões observados. Essa estratégia facilitou a prova de conceito, porém limitava reprodutibilidade.

O Berçário Sintético-Híbrido Governado responde a essa limitação. Em vez de quatro perfis construídos diretamente pelo pesquisador, ele define um espaço de 861 combinações com genealogia explícita. O modelo deixa de decidir livremente quais atributos caracterizam o respondente e passa a receber uma representação que já possui procedência determinada.

Essa expansão não deve ser interpretada como aumento de N empírico. Os 861 candidatos derivam das mesmas 42 entrevistas. A contribuição está em explorar sistematicamente combinações plausíveis e selecionar agentes de maneira menos dependente da intuição do pesquisador.

A diferença fica concreta quando se observa o que cada abordagem permite. Na construção artesanal, cabia ao pesquisador decidir quais atributos comporiam o respondente, inclusive atributos sem correspondência em qualquer entrevista do corpus. O berçário governado não admite essa operação: como todo candidato deriva do cruzamento de dois entrevistados reais e carrega genealogia registrada, nenhum agente pode exibir atributo ausente da base. A consequência é desconfortável e correta. O protocolo não corrige as ausências do corpus, torna-as visíveis, ao impedir que a recombinação simule uma diversidade que a coleta não produziu.

A avaliação formativa produziu, nesse sentido, um resultado próprio. Cada falha observada na demonstração antecedente foi convertida em um mecanismo específico do protocolo, associado a uma etapa do Quadro 1. O Quadro 3 registra esse mapeamento e torna auditável a passagem entre o que falhou e o que o artefato passou a controlar.

Falha observada na demonstração	**Mecanismo de controle incorporado**	**Etapa**
Contaminação entre falas do entrevistador e do participante	Marcação manual [E] e regra determinística de exclusão	5
Tratamento de 43 arquivos como 43 respondentes	Auditoria do corpus e consolidação em 42 unidades	2
Escores atribuídos sem aferição de estabilidade	Concordância entre codificadores, teste de sementes e auditoria de evidência	6 e 7
Perfis artesanais irreproduzíveis e dependentes da intuição	Crossover por blocos com semente fixa e genealogia registrada	8
Higienização não detectada na leitura da resposta	Verificação de ancoragem em etapa independente da geração	13
Perguntas de impacto material submetidas ao canal sintético	Gate de admissibilidade anterior à geração	0
Ausência de registro das condições de execução	Trilha obrigatória de serviço, modelo, versão, prompt e resposta bruta	15

Quadro 3: Da avaliação formativa aos controles do protocolo

Fonte: elaborado pelos autores.

A mudança também altera o papel da clusterização. Perfis deixam de ser definidos antecipadamente como “jovem digital”, “veterano cauteloso” ou categorias equivalentes. A solução de clusters deve emergir da matriz revisada. Os rótulos são atribuídos depois, a partir dos padrões encontrados.

4.4 A tipologia epistemológica como mecanismo de governança

HIPÓTESE, INDÍCIO e SINAL DE RISCO constituem a contribuição central do protocolo porque transformam uma classificação analítica em regra de decisão.

Uma HIPÓTESE pode entrar na próxima rodada de ideação ou ser convertida em pergunta para campo. Um INDÍCIO pode justificar a priorização de um teste, porque já possui apoio recorrente no corpus. Um SINAL DE RISCO interrompe o fluxo sintético e exige inspeção ou participação humana antes de qualquer decisão com consequência material.

O estatuto epistêmico também reduz um problema organizacional: a transformação gradual de outputs em "verdades" à medida que circulam entre documentos e reuniões. O registro obrigatório da categoria mantém a origem da afirmação visível durante o ciclo do projeto.

Quatro princípios de design organizam o artefato e podem ser transferidos para outros contextos de simulação em discovery. O primeiro é a genealogia obrigatória: nenhum agente entra em uso sem procedência registrada até os respondentes de origem, o que impede que atributos ausentes do corpus ingressem na simulação. O segundo é a separação entre gerador e juiz: quem produz a resposta não decide sobre sua validade, e a verificação de ancoragem ocorre em etapa independente. O terceiro é o gate anterior à geração: a decisão sobre o que pode ser simulado precede a simulação, em vez de ser negociada diante de uma resposta já formulada. O quarto é o estatuto epistêmico persistente: toda saída circula rotulada, e o rótulo acompanha a afirmação enquanto ela transita entre documentos e reuniões.

Nos termos de Gregor e Hevner (2013), a contribuição é de exaptação. Nem invenção, porque nenhum dos componentes é novo isoladamente, nem simples melhoria de solução conhecida no mesmo domínio. O que se propõe é a transferência de recursos consolidados em outros campos, recombinação por blocos com genealogia rastreável e controle de plausibilidade multivariada, para um domínio no qual não eram utilizados: a governança epistemológica do discovery qualitativo assistido por modelos de linguagem. O valor da exaptação não está nos componentes, e sim na arquitetura de controle que os torna auditáveis em um contexto de decisão com consequência material.

4.5 Gate 0: nem toda pergunta deve ser simulada

O protocolo acrescenta governança antes da geração, porque algumas perguntas não se tornam legítimas apenas porque o modelo consegue respondê-las. O Gate 0 diferencia decisões exploratórias de decisões materiais. Perguntas como "que objeções esse conceito pode provocar?" ou "qual condição do produto parece pouco clara?" podem ser submetidas ao berçário como preparação.

Questões como "qual taxa de juros o entregador aceitaria?", "quanto podemos cobrar?", "quem deve ser reprovado?" ou "qual regra de atraso maximiza recuperação?" possuem impacto direto e não devem ser decididas por resposta sintética. A fronteira importa especialmente em produtos financeiros, onde a assimetria de consequência é elevada: uma interpretação errada pode produzir endividamento, exclusão ou redução de liquidez para quem já opera sob renda instável.

4.6 Contribuição para Gestão de Projetos

O berçário modifica a forma como o gestor organiza aprendizado em projetos de produto. O Product Manager passa a trabalhar com dois circuitos de evidência, um sintético e outro orgânico, e precisa manter a separação entre eles.

Na prática, um ciclo pode começar com hipóteses originadas no corpus e nos princípios de design. Agentes selecionados tensionam essas hipóteses. Respostas classificadas como HIPÓTESE indicam novas perguntas. INDÍCIOS ajudam a priorizar quais pontos merecem prototipação. SINAIS DE RISCO definem onde a equipe não deve avançar sem contato humano.

O benefício gerencial não é "pesquisar sem usuários", e sim chegar ao usuário com um problema mais bem delimitado, o que altera a alocação de tempo da equipe e do participante.

O artefato também cria rastreabilidade de decisão: se uma funcionalidade for priorizada, o projeto registra se sua origem foi entrevista, princípio de design, resposta sintética ou validação de campo. Essa procedência é relevante em projetos com efeitos sociais porque torna auditável o caminho entre evidência e decisão.

4.7 Humanização do discovery e custo de oportunidade

A dimensão social do protocolo decorre de uma característica concreta do público pesquisado. Trabalhadores remunerados por entrega convertem tempo disponível em possibilidade de renda. A participação em pesquisa pode ter custo de oportunidade maior do que em populações cuja remuneração não depende diretamente do período trabalhado.

Eliminar entrevistas para reduzir esse custo contradiria o próprio argumento do artigo, porque tornaria a experiência humana dispensável justamente onde ela é necessária. O berçário adota outro princípio: reduzir contato de baixo valor e preservar contato de alto valor. Protótipos inconsistentes, hipóteses mal formuladas e perguntas refináveis antes do campo passam primeiro pela camada sintética. Produtos, preços, riscos, dados pessoais e consequências materiais retornam obrigatoriamente ao canal humano.

Humanização, nesse contexto, não significa maximizar o número de interações com pessoas. Significa tratar o tempo solicitado ao participante como recurso que precisa ser justificado pelo valor informacional da interação.

4.8 Limitações e avaliação subsequente

O protocolo possui limites que não são corrigidos pela quantidade de agentes gerados. O primeiro é a dependência do corpus. Ausências e sub-representações originais não desaparecem pelo crossover. O caso mais evidente é o de gênero: as 43 entrevistas foram realizadas apenas com homens, e nenhuma recombinação entre elas produz experiência feminina legítima. Os achados deste estudo referem-se, portanto, a entregadores homens, e essa fronteira deve acompanhar qualquer uso posterior dos princípios de design derivados do corpus.

A limitação mais importante, no entanto, é de estatuto. O berçário revisado ainda não foi executado sobre as 42 entrevistas. O que este artigo entrega é o design do artefato e a avaliação formativa que o originou, não evidência de desempenho. Nenhuma taxa de retenção, distribuição de distâncias ou solução de agrupamento descrita na seção 3 deve ser lida como resultado obtido. Registrar isso de forma direta parece preferível a sugerir, pela ordem da exposição, uma validação que não ocorreu.

O segundo limite é a transformação de uma análise qualitativa em representação escalar. As 25 dimensões são instrumentos computacionais, não substitutos da interpretação reflexiva. O corpus original permanece necessário para auditar os escores.

Há também dependência do modelo. LLMs diferentes podem responder de modo distinto ao mesmo perfil, e atualizações do fornecedor podem alterar resultados. Por isso, versão, prompt, data e parâmetros fazem parte da trilha de auditoria.

A avaliação empírica subsequente deverá confrontar agentes e entregadores reais com os mesmos protótipos e roteiro nuclear. O objetivo não será demonstrar equivalência global, mas identificar em que dimensões ocorre convergência, onde surgem divergências e quais tipos de conteúdo são sistematicamente higienizados ou omitidos.

O desenho previsto inclui nova amostra de 15 a 20 entregadores, com diversificação intencional de gênero, experiência, região e relação com plataformas. A ampliação não é refinamento amostral, e sim correção de ausências que a demonstração antecedente tornou evidentes. Enquanto essa coleta não ocorrer, o protocolo trata como SINAL DE RISCO, por definição, qualquer saída sintética que dependa de atributos ausentes do corpus, em vez de admiti-la como hipótese investigável por simulação.

5 CONCLUSÕES

Este artigo apresentou o design do Berçário Sintético-Híbrido Governado, infraestrutura metodológica para Product Discovery sob escassez algorítmica. O artefato parte de 43 arquivos documentais consolidados em 42 entrevistas independentes e define uma amostra reprodutível de 861 agentes candidatos. Cada candidato preserva genealogia e blocos derivados do corpus, passa por controle de plausibilidade auditado e só então pode ser selecionado para simulação. A contribuição reivindicada é de design, e a avaliação de desempenho pertence ao estudo seguinte.

A contribuição do método não reside no número de agentes. Multiplicar combinações derivadas do mesmo corpus não multiplica participantes. O ganho está na capacidade de explorar sistematicamente tensões já observadas, documentar a origem de cada representação e impedir que a fluência de uma resposta se converta automaticamente em insight de usuário.

A avaliação formativa forneceu o problema que o protocolo precisou resolver. Entre 12 respostas sintéticas, cinco foram classificadas como INDÍCIO, três como HIPÓTESE e quatro como SINAL DE RISCO, e surgiram lacunas de corpus e higienizações. Esses achados não validam a população de 861 agentes. Mostram por que uma camada explícita de governança, com aferição prévia de confiabilidade, é necessária antes de qualquer geração em escala.

A tipologia HIPÓTESE, INDÍCIO e SINAL DE RISCO organiza essa governança no nível das saídas. O Gate 0 atua antes da geração e delimita questões que não devem depender do sintético. A verificação de ancoragem atua depois, confrontando a resposta com o material orgânico. Entre esses dois controles, o agente pode funcionar como instrumento de triagem.

Para Gestão de Projetos, o protocolo atribui ao Product Manager uma responsabilidade adicional: governar a procedência da evidência utilizada no discovery. A simulação pode antecipar objeções, melhorar protótipos e preparar roteiros. Não deve decidir preço, risco, elegibilidade ou outras condições materiais sem participação humana.

A contribuição social deriva da mesma arquitetura. O tempo de trabalhadores submetidos à lógica do giro não deve ser tratado como insumo gratuito e ilimitado de pesquisa. Concentrar a participação humana onde ela é epistemologicamente necessária permite combinar tecnologia e centralidade das pessoas sem confundir uma com a outra.

A próxima avaliação deverá testar essa fronteira empiricamente. A comparação entre agentes e entregadores reais permitirá identificar regiões de convergência, divergência e higienização e recalibrar os gates do protocolo. Um resultado cientificamente útil não será demonstrar que o sintético "acerta" como uma pessoa. Será especificar, com evidência, em que circunstâncias ele ajuda o projeto a formular perguntas melhores e em que momento deve deixar de falar.

REFERÊNCIAS

Argyle, L. P., Busby, E. C., Fulda, N., Gubler, J., Rytting, C., & Wingate, D. (2023). Out of one, many: Using language models to simulate human samples. Political Analysis, 31(3), 337–351. https://doi.org/10.1017/pan.2023.2

Autor. (2026). Referência omitida para avaliação cega.

Bisbee, J., Clinton, J. D., Dorff, C., Kenkel, B., & Larson, J. M. (2024). Synthetic replacements for human survey data? The perils of large language models. Political Analysis, 32(4), 401–416. https://doi.org/10.1017/pan.2024.5

Braun, V., & Clarke, V. (2021). One size fits all? What counts as quality practice in (reflexive) thematic analysis? Qualitative Research in Psychology, 18(3), 328–352. https://doi.org/10.1080/14780887.2020.1769238

Camuffo, A., Cordova, A., Gambardella, A., & Spina, C. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial. Management Science, 66(2), 564–586. https://doi.org/10.1287/mnsc.2018.3249

Dillion, D., Tandon, N., Gu, Y., & Gray, K. (2023). Can AI language models replace human participants? Trends in Cognitive Sciences, 27(7), 597–600. https://doi.org/10.1016/j.tics.2023.04.008

Gilardi, F., Alizadeh, M., & Kubli, M. (2023). ChatGPT outperforms crowd workers for text-annotation tasks. Proceedings of the National Academy of Sciences, 120(30), Article e2305016120. https://doi.org/10.1073/pnas.2305016120

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. MIS Quarterly, 37(2), 337–355. https://doi.org/10.25300/MISQ/2013/37.2.01

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. MIS Quarterly, 28(1), 75–105. https://doi.org/10.2307/25148625

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. Academy of Management Annals, 14(1), 366–410. https://doi.org/10.5465/annals.2018.0174

Lindgren, E., & Münch, J. (2016). Raising the odds of success: The current state of experimentation in product development. Information and Software Technology, 77, 80–91. https://doi.org/10.1016/j.infsof.2016.04.008

Mani, A., Mullainathan, S., Shafir, E., & Zhao, J. (2013). Poverty impedes cognitive function. Science, 341(6149), 976–980. https://doi.org/10.1126/science.1238041

Messeri, L., & Crockett, M. J. (2024). Artificial intelligence and illusions of understanding in scientific research. Nature, 627(8002), 49–58. https://doi.org/10.1038/s41586-024-07146-0

O'Connor, C., & Joffe, H. (2020). Intercoder reliability in qualitative research: Debates and practical guidelines. International Journal of Qualitative Methods, 19, 1–13. https://doi.org/10.1177/1609406919899220

Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. Journal of Management Information Systems, 24(3), 45–77. https://doi.org/10.2753/MIS0742-1222240302

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. International Journal of Communication, 10, 3758–3784. https://ijoc.org/index.php/ijoc/article/view/4892

Venable, J., Pries-Heje, J., & Baskerville, R. (2016). FEDS: A framework for evaluation in design science research. European Journal of Information Systems, 25(1), 77–89. https://doi.org/10.1057/ejis.2014.36

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. Work, Employment and Society, 33(1), 56–75. https://doi.org/10.1177/0950017018785616