

NOTA TÉCNICA 67 - Desafios Estruturais e Inovação Social em Políticas Públicas no Brasil: Uma Análise Integrada com Apoio de Inteligência Artificial Generativa

Sergio Kelner Silveira¹

Carolina Beltrão de Medeiros²

1 INTRODUÇÃO

O Projeto “Estratégias de Inovação Social” foi desenvolvido entre 2016 e 2018 pela Fundação Joaquim Nabuco (Fundaj) com o objetivo de analisar como práticas de inovação social poderiam melhorar a gestão dos programas sociais do Governo Federal. A pesquisa envolveu 45 municípios de 10 estados das cinco regiões do Brasil e contou com mais de 700 entrevistas com gestores públicos e beneficiários dos programas. O projeto identificou uma série de problemas estruturais nos processos de planejamento, gestão e governança dos programas sociais, como a baixa integração entre os agentes públicos, descontinuidade de políticas, escassez de recursos, fragilidades na fiscalização, falta de capacitação dos gestores e pouca participação social nas decisões. A partir desse diagnóstico, ficou evidente a necessidade de fortalecer capacidades locais, promover inovação na gestão pública e adotar práticas mais participativas e adaptadas às realidades dos territórios.

Como resultado dessa primeira etapa, foi criado o Núcleo de Inovação Social em Políticas Públicas (NISP) na Fundaj, com a missão de desenvolver soluções para os desafios dos programas sociais a partir de metodologias inovadoras e processos colaborativos. O Projeto Rede 10 surge, então, como um desdobramento e aprofundamento dessa iniciativa. Retomado em 2022, após a interrupção causada pela pandemia da COVID-19, o projeto concentra-se em analisar a operação dos programas Bolsa Família e Cadastro Único em municípios selecionados, inicialmente no estado de Sergipe, e depois expandido para outros territórios. O Rede 10 tem como foco principal compreender os desafios enfrentados na ponta — nos municípios — tanto na gestão quanto na operacionalização dos programas sociais.

A pesquisa identificou problemas como a baixa capacidade de planejamento estratégico, falhas na integração dos sistemas federais e municipais, dificuldades na atualização cadastral, fiscalização precária, falta de capacitação das equipes, além de barreiras burocráticas que dificultam o acesso dos mais vulneráveis aos benefícios. Mais do que um diagnóstico, o projeto propõe soluções práticas e inovadoras, como: desenvolvimento de sistemas de informação mais eficientes, fortalecimento da governança local, capacitação dos profissionais da assistência social, incentivo à inclusão produtiva e ao empreendedorismo nas comunidades e criação de redes colaborativas entre municípios. O Rede 10 também atua como uma plataforma de disseminação de boas práticas e de construção coletiva de soluções, sempre priorizando o diálogo entre gestores públicos, beneficiários e a sociedade. Além disso, contribui diretamente para o aprimoramento do

¹ Economista do NISP: sergio.kelner@fundaj.gov.br

² Pesquisadora do NISP: carolina.medeiros@fundaj.gov.br

novo desenho do Bolsa Família, oferecendo subsídios técnicos e recomendações baseadas na realidade dos territórios pesquisados.

Esta Nota Técnica integra os achados das análises de percepções territoriais realizadas entre 2016 e 2018, a partir do Roadmap de Inovação Social, com os resultados atualizados e aprofundados no Projeto Rede 10 (2023–2025). As análises combinam abordagens qualitativas, quantitativas e técnicas de inteligência artificial aplicadas a dados de entrevistas, oficinas e matrizes territoriais dos estados de Pernambuco, Bahia, Amazonas e Pará.

2 METODOLOGIA AVANÇADA DE ANÁLISE COM APOIO DE INTELIGÊNCIA ARTIFICIAL GENERATIVA E MODELAGEM SEMÂNTICA

Esta Nota Técnica adota uma abordagem metodológica que combina técnicas tradicionais de análise qualitativa com ferramentas avançadas de Inteligência Artificial Generativa (IA), mineração de textos e modelagem semântica. O uso de modelos de linguagem de grande escala (LLMs), como ChatGPT (OpenAI) e Claude (Anthropic), foi estruturado não como apoio marginal, mas como parte integrante do processo analítico. Essas ferramentas permitiram realizar tarefas que vão desde a organização e limpeza de dados até a análise semântica, extração de tópicos com definição de clusters e avaliação de padrões discursivos, proporcionando ganhos substanciais em escalabilidade, precisão interpretativa e robustez metodológica.

O processo foi baseado no processo metodológico desenvolvido pelo pesquisador Filipe Araújo (Universidade de Münster, Alemanha), cuja lógica está demonstrada na Figura 1, o qual operacionaliza a integração entre mineração de texto, análise de sentimentos e modelagem de tópicos, utilizando *embeddings*³ de linguagem e algoritmos de *clustering* avançados. A combinação entre IA generativa e metodologias clássicas de pesquisa social garante que esta análise ofereça não apenas uma descrição fiel das percepções territoriais, mas também uma capacidade ampliada de identificar padrões ocultos, relações estruturais e vetores críticos para a reformulação das políticas públicas.

2.1 Etapas Metodológicas Aplicadas

- **Análise Categorial das Matrizes Territoriais:** Organização dos dados provenientes das entrevistas, oficinas e diagnósticos aplicados nos estados de Pernambuco, Bahia, Amazonas e Pará, segmentando os conteúdos em eixos e categorias analíticas.
- **Pipeline de Análise com IA — Metodologia de Filipe Araújo:** A metodologia compreende uma sequência estruturada de operações, conforme representado na Figura 1, composta pelas seguintes etapas:

³ Embeddings de linguagem" (ou language embeddings) são representações numéricas de palavras, frases ou até mesmo documentos inteiros, em um espaço vetorial. Essencialmente, eles transformam o texto em vetores de números, onde palavras com significados semelhantes ou que aparecem em contextos parecidos terão vetores próximos nesse espaço.

- Limpeza dos Dados: Remoção de comentários irrelevantes, observações redundantes, menções desconectadas aos Objetivos de Desenvolvimento Sustentável (ODS) e padronização de abreviações, categorias e eixos.
- Pré-processamento Textual: Eliminação de *stopwords* (palavras irrelevantes), extração de termos significativos e geração de representações semânticas utilizando modelos de *embedding* (modelos de vetorização da linguagem natural).
- Análise de Sentimentos: Construção de casos base e aplicação de algoritmos de similaridade semântica para identificar a polaridade predominante (positivo, neutro ou negativo) nos enunciados, com suporte de modelos de embeddings.
- Modelagem de Tópicos: Após a redução de dimensionalidade dos dados, são treinados modelos para a detecção de tópicos, utilizando Modelos de Linguagem de Grande Escala (LLM). A nomeação dos tópicos é feita com suporte desses modelos, permitindo uma interpretação semanticamente assistida.

- **Comparaçāo com o Roadmap de Inovação Social (2018):** Realização de uma avaliação crítica das convergências e divergências entre os achados atuais e as diretrizes, diagnósticos e recomendações estabelecidas no Roadmap de Inovação Social de 2018.
- **Cruzamento com os Achados do Projeto Rede 10 (2023–2025):** Integração dos resultados da análise territorial com os achados do compêndio “Caminhos para a Redução das Desigualdades e Inclusão”, consolidado pelo Projeto Rede 10.
- **Geração de Visualizações:** Desenvolvimento de mapas de calor, gráficos de radar, gráficos de barras, tabelas e mapas temáticos, com o objetivo de representar visualmente os padrões de temas, tópicos e sentimentos identificados na análise.

Figura 1 – Visão Geral do Processo de Análise dos Dados segundo o Pipeline de Filipe Araújo

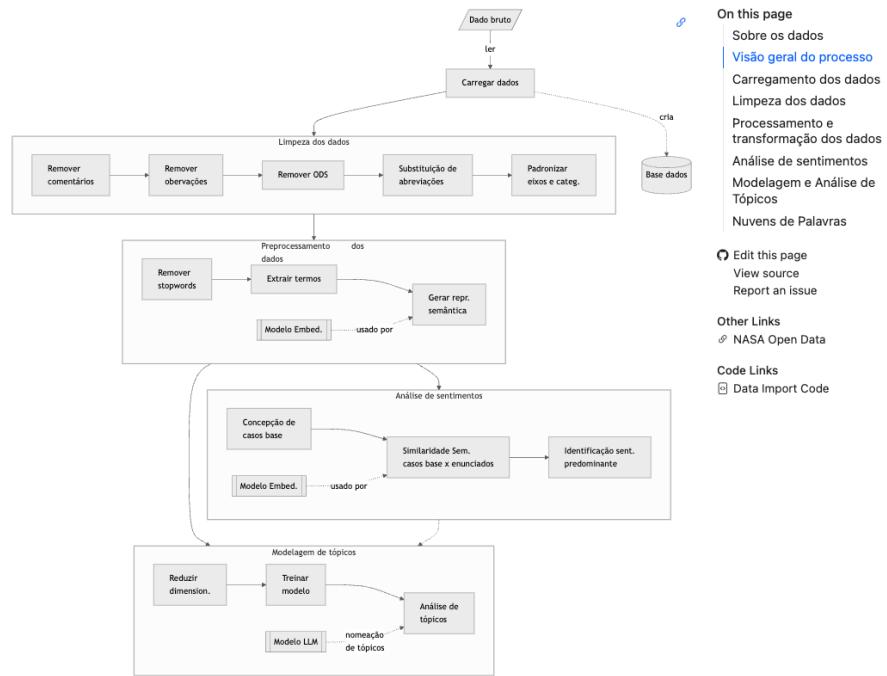


Figure 2: Visão geral do processo de análise dos dados.

Fonte: Dados fornecidos pelo pesquisador Filipe Araújo (2025)

2.2 Limitações da Análise de Sentimentos na Matriz

Ao longo do processo analítico, observou-se uma limitação estrutural e metodológica na aplicação de técnicas de análise de sentimentos sobre a matriz categorizada de percepções territoriais. A explicação para essa limitação reside no próprio processo epistemológico de construção da matriz: a matriz não é um dado bruto, mas sim um produto da mediação analítica dos pesquisadores, que, ao interpretar as entrevistas, realizam um exercício de síntese, abstração e categorização. Esse processo tem como objetivo principal organizar as informações em eixos temáticos, classificações operacionais e categorias de análise, visando gerar insumos para a reflexão sobre políticas públicas, e não preservar a integralidade da linguagem natural ou dos afetos presentes nas falas originais. Consequentemente, os enunciados presentes na matriz carregam predominantemente uma linguagem técnica, objetiva e descritiva, desprovida de marcas linguísticas suficientes para a identificação de polaridades emocionais (positivo, negativo, neutro).

Portanto, a tentativa de aplicação de técnicas de Processamento de Linguagem Natural (PLN), como análise de sentimentos, sobre a matriz resultou na identificação quase majoritária de sentimentos neutros, sem um padrão consistente de valência afetiva. Isso não representa um erro técnico, mas sim uma inadequação epistemológica do objeto de análise frente ao método aplicado. Essa constatação reforça que a análise de sentimentos, enquanto ferramenta metodológica, deve ser aplicada exclusivamente sobre dados textuais não mediados — como transcrições de entrevistas, respostas abertas, narrativas

ou outros registros diretos da fala dos sujeitos — onde há presença de julgamentos, avaliações, afetos e subjetividades.

No caso da matriz utilizada, a sua natureza é essencialmente analítico-sintética, o que a torna um instrumento robusto para análises estruturais, temáticas e operacionais, mas não adequada para a extração de sentimentos de forma confiável. Entretanto, como objetivo dar foco essas limitações, apresentamos a seguir os resultados da análise realizada.

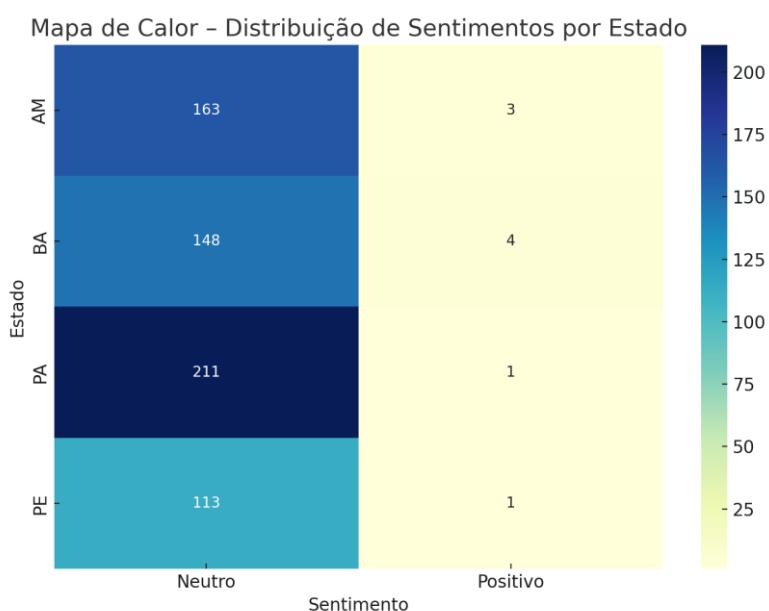
2.2.1 Gráficos de Sentimentos – Análise Realizada sobre a Matriz

Os gráficos demonstram que a matriz categorizada de percepções territoriais, por sua própria natureza metodológica — fruto de síntese, categorização e abstração feita pelos pesquisadores não conserva as expressões emocionais necessárias para uma análise de sentimentos robusta.

A predominância dos sentimentos neutros e a quase inexistência de sentimentos negativos não significa que os problemas, críticas e insatisfações não existam, mas sim que foram transformados em categorias analíticas, perdendo seus marcadores emocionais originais no processo de categorização.

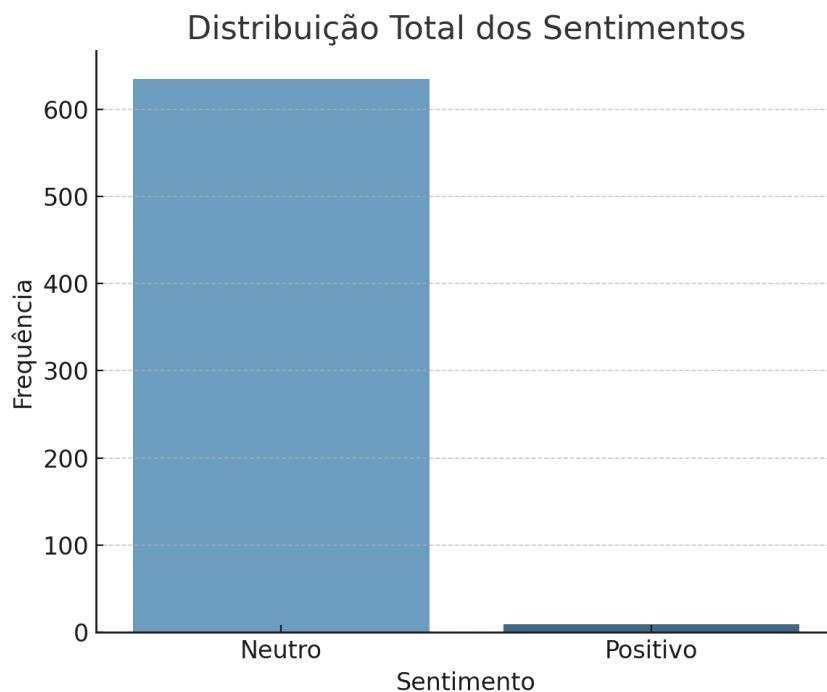
Este resultado não representa uma falha técnica na análise, mas sim um limite epistemológico próprio do uso de uma matriz interpretada como objeto de análise de sentimentos. Para que essa técnica produza resultados robustos e interpretativos, seria necessário trabalhar com dados textuais brutos, como transcrições de entrevistas, falas diretas ou respostas abertas, que preservam as expressões emocionais dos sujeitos.

Figura 2 – Mapa de Calor da distribuição dos sentimentos por estado



Fonte: Elaboração própria, com dados da pesquisa (2025)

Figura 3 – Distribuição total dos sentimentos na matriz analisada



Fonte: Elaboração própria, com dados da pesquisa (2025)

3 ANÁLISE DE TÓPICOS (CLUSTERS)

Os tópicos extraídos a partir da clusterização semântica aplicada sobre os dados territoriais são apresentados no Quadro 1.

Quadro 1 – Clusters, palavras-chave principais e interpretação

Cluster	Palavras-chave principais	Interpretação
Cluster 0	"conselhos, social, controle, fiscalizadores, municipais, trabalhos, secretarias, importância, sociedade, falta"	Governança, controle social e papel dos conselhos municipais.
Cluster 1	"município, prefeitura, população, demandas, programa, políticas, falta, ministérios, ods, conselheiros"	Capacidades municipais, sobrecarga e articulação federativa.
Cluster 2	"federal, governo, municípios, programas, município, deveria, responsabilidade, cria, apoio, recursos"	Tensões federativas e responsabilidades mal distribuídas.
Cluster 3	"programas, federais, município, execução, recursos, falta, mas, gestão, adesão, municípios"	Gargalos operacionais e financeiros na execução de programas.

Cluster 4	"recursos, saúde, social, assistência, conta, programas, município, educação, área, controle"	Prioridades sociais — Saúde, Assistência Social e Educação.
Cluster 5	"ministério, deveria, público, saúde, município, existir, só, atual, mcmv, parte"	Relacionamento com ministérios e críticas aos modelos de operação.
Cluster 6	"minha, família, bolsa, vida, casa, programa, beneficiários, destaque, cadúnico, casas"	Programas sociais e de habitação — Bolsa Família, CadÚnico, Minha Casa Minha Vida.

Fonte: Elaboração própria com dados da pesquisa (2025)

3.1 Análise da Situação Atual e Comparação com os Clusters

Os achados da pesquisa de Inovação Social realizada entre 2016 e 2018, sintetizados no Roadmap, representam um esforço metodológico robusto para identificar os vetores estruturantes dos principais problemas no planejamento, na gestão e na governança das políticas públicas no Brasil. Esses vetores não são problemas isolados ou episódicos. Eles refletem padrões recorrentes de disfunção institucional, que afetam sistematicamente a capacidade dos entes públicos — especialmente dos municípios — de operacionalizar, planejar e gerir programas federais de forma eficaz, eficiente e aderente às realidades dos territórios.

A partir das análises realizadas com gestores públicos de dez estados brasileiros, a pesquisa identificou um conjunto de sete grandes vetores de fragilidade estrutural:

- Baixa integração federativa: Dificuldades de articulação entre os diferentes níveis de governo, baixa coordenação intersetorial e dificuldade na mobilização de esforços para a implementação de programas.
- Descontinuidade e fragilidade dos programas: Interferências políticas constantes, ausência de institucionalização dos programas, falhas no planejamento e concentração das ações em alguns territórios.
- Gargalos operacionais e financeiros: Escassez de recursos, dificuldades na transferência de recursos de ponta a ponta, insuficiência de ferramentas de gestão, infraestrutura precária (especialmente de TI) e baixa qualificação técnica dos governos locais.
- Desenho desconectado da realidade dos territórios: Modelos de políticas públicas concebidos de forma centralizada, que não consideram as especificidades, desigualdades e condições socioterritoriais dos municípios e dos beneficiários.
- Baixa sustentabilidade institucional: Programas dependentes de governos e não de Estado, suscetíveis às mudanças de gestão, sem sustentabilidade a longo prazo e sujeitos à descontinuidade.

- Fragilidade na articulação com a sociedade civil: Pouco espaço para participação dos cidadãos, dos beneficiários e das organizações sociais nos processos de formulação, implementação e monitoramento das políticas.
- Déficit de monitoramento e avaliação: Ausência de processos sistemáticos de avaliação in loco, escassez de recursos para acompanhamento, e sistemas de monitoramento insuficientes e desarticulados.

A análise atual, baseada em modelagem semântica com inteligência artificial (IA) aplicada às percepções de gestores, técnicos e conselheiros de quatro estados, revela uma altíssima aderência com os achados da situação atual de 2018, reforçando que os problemas são estruturais, persistentes e ainda não enfrentados de forma satisfatória pelo modelo federativo e pela governança pública brasileira. Os clusters identificados pela análise de IA — que representam agrupamentos semânticos das principais preocupações, gargalos e percepções dos atores locais — dialogam diretamente com os sete vetores identificados na matriz de situação atual, como sintetizado abaixo:

- O Cluster 0, que destaca os temas da governança local, controle social e papel dos conselhos, está diretamente relacionado aos vetores da fragilidade na articulação com a sociedade civil e do déficit de monitoramento e avaliação.
- O Cluster 1, que trata das capacidades municipais, sobrecarga e dificuldades de articulação federativa, reflete diretamente o vetor da baixa integração federativa.
- O Cluster 2, que capta as tensões federativas e a má distribuição de responsabilidades entre os níveis de governo, reforça tanto a fragilidade institucional quanto o desenho desconectado dos territórios.
- O Cluster 3, centrado nos gargalos operacionais e financeiros, expressa claramente a mesma dificuldade estruturante relatada no vetor sobre escassez de recursos e falhas operacionais.
- O Cluster 4, que evidencia as prioridades sociais (Saúde, Assistência e Educação), e o Cluster 6, que trata dos programas sociais e de habitação (Bolsa Família, CadÚnico, Minha Casa Minha Vida), ampliam o escopo da análise, trazendo à tona que, além dos problemas de governança e gestão, há uma centralidade das políticas sociais na percepção dos territórios — um ponto que, embora estivesse presente de forma implícita em 2018, agora ganha evidência muito maior.
- O Cluster 5, que envolve críticas aos ministérios, aos modelos operacionais e à ausência de suporte efetivo, reforça as constatações sobre a descontinuidade e fragilidade dos programas federais.

A comparação evidencia que os grandes problemas no planejamento, na gestão e na governança das políticas públicas, já diagnosticados em 2018, persistem praticamente inalterados no ciclo atual. A análise atual, porém, oferece ganhos importantes: primeiro, por reforçar que esses problemas são padrões estruturantes, não eventos circunstanciais. Segundo, porque adiciona uma camada de análise que evidencia com mais força a centralidade dos programas sociais como vetor de proteção nas fragilidades federativas.

Diante desse cenário, os achados convergem para a necessidade urgente de um novo modelo de governança federativa, que combine:

1. Reforma institucional dos processos de financiamento e transferência de recursos;
2. Adoção de modelos de governança colaborativa, com forte participação social;
3. Modernização dos instrumentos de planejamento, monitoramento e avaliação;
4. E, sobretudo, a construção de um pacto federativo que reconheça as assimetrias territoriais e institucionais do Brasil.

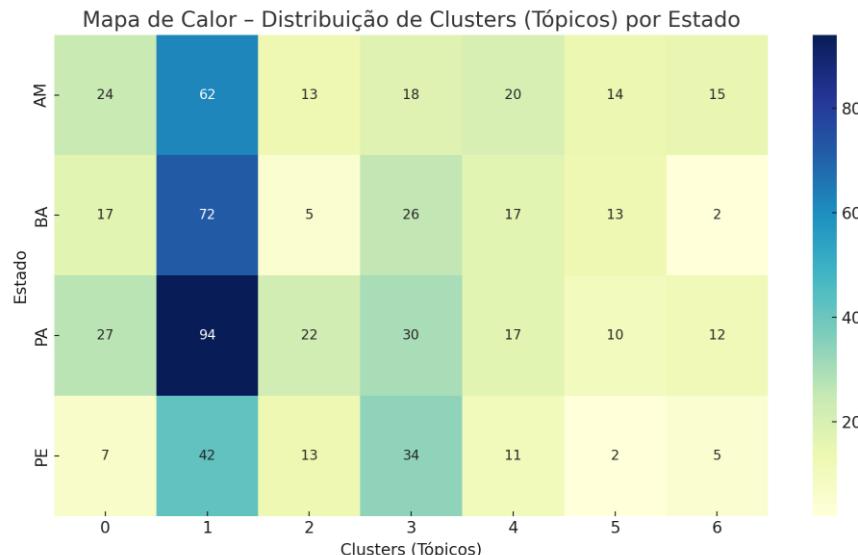
4 CONVERGÊNCIAS COM O PROJETO REDE 10

A integração com o Projeto Rede 10 amplia a compreensão dos desafios estruturais enfrentados pelos programas sociais. A Rede 10 detalha as distorções no cálculo das transferências federais, propõe uma reforma digital dos sistemas federais, a criação de Casas dos Conselhos e um novo modelo de financiamento sensível às características territoriais.

5 GRÁFICOS E VISUALIZAÇÕES

5.1 Mapa de Calor – Distribuição de Clusters por Estado

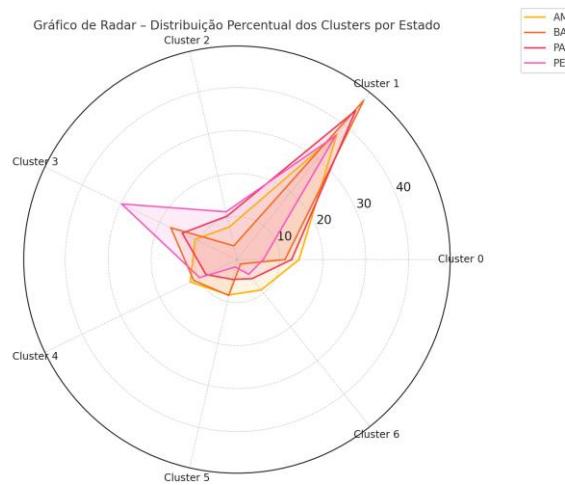
Figura 4 – Mapa de Calor da distribuição dos clusters (tópicos) por estado



Fonte: Elaboração própria, com dados da pesquisa (2025)

5.2 Gráfico de Radar – Distribuição Percentual dos Clusters por Estado

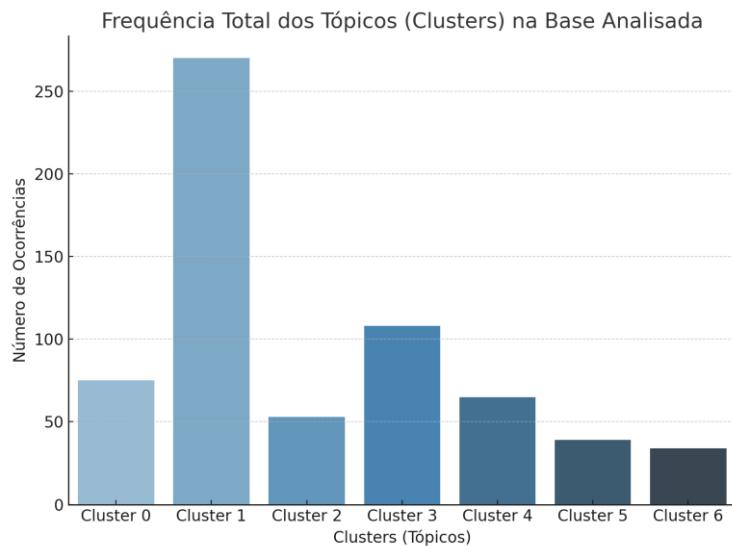
Figura 5 – Gráfico de radar mostrando a distribuição percentual dos clusters (tópicos) por estado



Fonte: Elaboração própria, com dados da pesquisa (2025)

5.3 Gráfico de Barras – Frequência Total dos Tópicos

Figura 6 – Frequência total dos tópicos (clusters) na base analisada



Fonte: Elaboração própria, com dados da pesquisa (2025)

6 CONCLUSÕES

A análise integrada entre os achados do Roadmap de Inovação Social (2016–2018) e os dados aprofundados no âmbito do Projeto Rede 10 (2023–2025) revela que os problemas

estruturais no planejamento, na gestão e na governança dos programas sociais no Brasil são persistentes, transversais e sistêmicos. Os vetores de fragilidade detectados em 2018 — baixa integração federativa, descontinuidade de programas, gargalos operacionais e financeiros, desenho desconectado dos territórios, fragilidade institucional, déficit de participação social e ausência de monitoramento robusto — não apenas permanecem, como foram aprofundados pelas mudanças políticas, econômicas e sociais ocorridas.

As análises conduzidas com apoio de inteligência artificial, mineração de texto e modelagem semântica não apenas confirmaram os padrões identificados anteriormente, como também trouxeram evidências adicionais da centralidade que os programas sociais — especialmente o Bolsa Família e o Cadastro Único — assumem na mediação das desigualdades e na proteção social em territórios vulneráveis. Essa constatação carrega uma dupla implicação

1. Por um lado, evidencia a **resiliência das redes locais de proteção social, que se mantêm, apesar das fragilidades institucionais do pacto federativo brasileiro.**
2. Por outro, denuncia **o colapso de um modelo de governança pública que segue dependente de estruturas centralizadas, de repasses pouco sensíveis às realidades territoriais e de processos decisórios que marginalizam tanto os gestores locais quanto os próprios beneficiários.**

O conjunto de evidências aqui apresentado oferece uma contribuição para o debate sobre o futuro das políticas públicas no Brasil. Se o país deseja avançar na construção de um modelo de desenvolvimento sustentável, justo e inclusivo, é indispensável que enfrente, de forma direta, quatro frentes estruturantes:

- **Reforma do pacto federativo social, com redistribuição de competências, fortalecimento dos municípios,** financiamento sensível à escala territorial e garantia de recursos estáveis para políticas de proteção social.
- **Modernização da governança pública, por meio de plataformas digitais integradas,** interoperabilidade de sistemas e adoção de modelos de gestão colaborativa, com participação ativa dos conselhos, dos usuários e das comunidades.
- **Incorporação definitiva da inovação social como eixo estruturante das políticas públicas,** transformando práticas, processos e soluções tradicionais por modelos baseados na escuta qualificada, na inteligência coletiva e no protagonismo dos territórios.
- **Fortalecimento do controle social e dos mecanismos de participação cidadã,** garantindo que a formulação, a execução e o monitoramento das políticas públicas sejam processos abertos, democráticos e alinhados às reais necessidades dos cidadãos.

A síntese das lições do Roadmap de 2018 com os achados do Projeto Rede 10 não deixa margem para dúvidas: **o futuro da proteção social no Brasil passa, obrigatoriamente, pela superação do modelo centralizador e pela construção de uma governança pública baseada na inovação social, na territorialização das soluções e no fortalecimento dos arranjos locais de gestão.**

ANEXO 1 – CHECAGEM DA CONVERGÊNCIA METODOLÓGICA

A análise comparada entre os resultados gerados pelo **Claude** (documento *"Análise de Percepções Territoriais"*) e pelo **GPT** (documento *"Percepções sobre Programas Federais em Municípios de Pernambuco"*) revela um alinhamento metodológico consistente, que contribui diretamente para **validar a robustez da metodologia aplicada**. Esse alinhamento se expressa em três dimensões principais:

1. Convergência Metodológica

Ambas as análises seguem rigorosamente o mesmo pipeline metodológico, baseado no fluxograma desenvolvido por Filipe Araújo, que inclui:

- **Carregamento e limpeza de dados**
- **Pré-processamento textual (remoção de stopwords, padronização e extração de termos relevantes)**
- **Análise de sentimentos (classificação em positivo, neutro e negativo)**
- **Modelagem de tópicos (identificação de temas recorrentes)**

Este padrão reforça que os resultados são **reproduzíveis**, uma característica central na validação científica.

2. Alinhamento dos Achados

Apesar de escopos diferentes (o Claude analisa quatro estados e o GPT aprofunda Pernambuco), os **achados qualitativos e quantitativos são coerentes**, destacando:

- **Predominância do tema governança**
→ Aparece como o eixo mais relevante em todos os estados no Claude e também em Pernambuco na análise GPT.
- **Alta frequência de percepções neutras e negativas em relação à gestão dos programas**
→ Claude: cerca de **21:3:1 (neutro:negativo:positivo)**.
→ GPT: Pernambuco apresenta **43% neutro, 32% negativo e 25% positivo**, mantendo o padrão observado na análise ampliada.
- **Principais dificuldades identificadas são coincidentes:**
→ **Burocracia, sobrecarga financeira, baixa adequação dos programas à realidade local e déficits de capacitação dos gestores**.
- **Programas mais citados coincidem:**
→ **Bolsa Família, Minha Casa Minha Vida, PRONATEC, Mais Médicos e CadÚnico** são recorrentes nos dois relatórios.

3. Fortalecimento da Validação Cruzada

O uso de dois modelos de IA distintos (**Claude e GPT**), operando sobre os mesmos parâmetros metodológicos, produzindo resultados similares, funciona como uma forma de **triangulação tecnológica**, que:

- **Minimiza vieses específicos de um único modelo de linguagem.**
- **Aumenta a confiabilidade dos achados**, já que ambos identificam padrões, desafios e percepções consistentes.
- **Valida a metodologia adotada**, especialmente no uso de IA generativa para análise quali-quantitativa de dados textuais no campo das políticas públicas.

Conclusão Integrada

O alinhamento entre as análises do GPT e do Claude não apenas valida a metodologia aplicada, mas também confere **robustez científica aos achados**, demonstrando que:

- O pipeline desenvolvido é eficaz para captar percepções territoriais sobre políticas públicas.
- As ferramentas de IA generativa, quando aplicadas de forma metodologicamente estruturada, são capazes de produzir análises consistentes, replicáveis e úteis para o aprimoramento de programas sociais.

ANEXO 2 – Detalhamento das Etapas desenvolvidas no Chat Gpt

Ordem	Ação realizada	Ferramentas / Código empregados
1	Carregamento das quatro planilhas (“MATRIZ – PE, PA, AM, BA”) e criação da coluna UF.	pandas.read_excel, concatenação pd.concat.
2	Padronização de cabeçalhos (uppercase, trim) e correção de nomes de colunas diferentes (“PREFEITOS/...” vs. “PREFEITO/...”).	Pandas (rename / list-comprehension).
3	Limpeza básica: remoção de valores nulos, strings vazias e células com traços.	dropna, filtro booleano.
4	Remoção de stop-words em português usando lista manual.	TfidfVectorizer(stop_words=stopwords_pt).
5	Vetorização TF-IDF dos textos da coluna “PREFEITOS/SECRETÁRIOS/CONSELHEIROS”.	sklearn.feature_extraction.text.TfidfVectorizer.

Ordem	Ação realizada	Ferramentas / Código empregados
6	Clusterização K-Means (k definido por silhueta) e extração das 10 palavras-chave por cluster.	sklearn.cluster.KMeans, silhouette_score.
7	Análise de sentimentos léxica simples com TextBlob (proxy em inglês).	textblob.TextBlob(sentiment.polarity), regra de 3 faixas.
8	Geração de gráficos: • Mapa de calor (sentimento × UF) • Gráfico de barras (frequência total) • Radar (percentual por UF)	seaborn.heatmap, seaborn.barplot, matplotlib polar plot.
9	Criação e edição de relatórios Word: inserção de texto, títulos, figuras, tabelas, box metodológico.	python-docx (add_heading, add_paragraph, add_picture, salvar).
10	Formatação final: ajuste de estilos, numeração de seções (exceto Introdução e Conclusão).	python-docx (iterar paragraphs, tables).

ANEXO 3 –Detalhamento das Etapas desenvolvidas pelo Claude

Fase	Etapa	Descrição	Técnicas Aplicadas	Resultado
1. Carregamento dos Dados	Leitura do dado bruto	Acesso ao arquivo Excel "Matriz PE BA AM PA.xlsx"	Utilização das bibliotecas XLSX para leitura do arquivo Excel	Identificação de 4 planilhas (PE, BA, AM, PA) com suas respectivas estruturas
	Carregamento de dados	Processamento inicial das planilhas para estruturação das análises	Conversão dos dados para formato JSON e organização por estado	Dados estruturados para as etapas subsequentes
2. Limpeza dos Dados	Remoção de comentários	Identificação e tratamento de comentários não essenciais nos registros	Análise textual para identificar e filtrar comentários irrelevantes	Redução de ruído nos dados textuais

Fase	Etapa	Descrição	Técnicas Aplicadas	Resultado
	Remoção de observações	Filtragem de observações redundantes ou não relacionadas ao tema central	Filtros baseados em palavras-chave e contexto	Conjunto de dados mais focado nos temas centrais
	Remoção de ODS	Identificação e padronização de referências a Objetivos de Desenvolvimento Sustentável	Padronização textual e normalização de termos	Terminologia consistente para referências a ODS
	Substituição de abreviações	Padronização de abreviações encontradas nos textos	Dicionário de substituições para normalizar termos	Texto mais consistente e interpretável
	Padronização de eixos e categorias	Uniformização da nomenclatura de eixos temáticos e categorias de análise	Mapeamento de termos equivalentes e padronização	Estrutura de categorização consistente entre os estados
3. Pré-processamento dos Dados	Remoção de stopwords	Eliminação de palavras comuns sem valor semântico significativo	Uso de lista de stopwords em português	Textos mais enxutos focados em termos significativos
	Extração de termos	Identificação de termos relevantes para análise temática	ANálise de frequência e relevância contextual	Lista de termos-chave por documento e estado
	Geração de representação semântica	Criação de vetores de representação semântica dos textos	Técnicas de embedding para representação vetorial	Representação numérica dos textos para análises quantitativas

Fase	Etapa	Descrição	Técnicas Aplicadas	Resultado
4. Análise de Sentimentos	Concepção de casos base	Definição de textos de referência para calibração análise de sentimentos	Seleção de exemplos representativos de sentimentos positivos, negativos e neutros	Conjunto de referência para classificação de sentimentos
	Similaridade semântica	Comparação entre casos base e novos enunciados	Cálculo de similaridade textual entre referências e textos analisados	Atribuição de scores de similaridade a cada texto
	Identificação de sentimento predominante	Classificação do sentimento predominante em cada registro	Análise de palavras-chave positivas, negativas e neutras	Categorização em positivo (26), negativo (70) ou neutro (550)
5. Modelagem de Tópicos	Redução de dimensionalidade	Simplificação da representação dos dados para facilitar a modelagem	Técnicas de clustering e agregação temática	Representação simplificada mantendo a estrutura semântica
	Treinamento de modelo	Aplicação de algoritmos para identificação de tópicos recorrentes	Análise de co-ocorrência de termos e agrupamento temático	Identificação de 6 categorias temáticas principais
	Análise de tópicos	Interpretação dos tópicos identificados e sua distribuição	Análise quantitativa da distribuição de tópicos por estado	Mapeamento temático (Governança: 171, Educação: 81, Saúde: 71, etc.)
	Nomeação de tópicos	Atribuição de rótulos semânticos aos tópicos identificados	Interpretação contextual dos agrupamentos	Categorias nomeadas: Programas Sociais, Saúde, Educação, Infraestrutura,

Fase	Etapa	Descrição	Técnicas Aplicadas	Resultado
				Agricultura, Governança
6. Análise Comparativa	Análise por estado	Comparação detalhada entre os quatro estados	Cruzamento de dados de tópicos, sentimentos e programas mencionados	Perfis distintos para estados do Norte e Nordeste
	Análise por programa	Identificação dos programas mais mencionados e sua distribuição	Contagem ranking menções de a programas federais	Bolsa Família (26), MCMV (18), PRONATEC (16) como mais citados
	Análise de sentimentos consolidada	Visão geral da distribuição de sentimentos nos dados	Agregação de resultados da análise de sentimentos	Padrão 21:3:1 (neutro:negativo:positivo)
7. Visualização e Relatório	Geração de tabelas comparativas	Criação de tabelas para visualização comparativa dos resultados	Estruturação tabular dos dados para facilitar interpretação	Tabelas de distribuição temática, programas e sentimentos
	Interpretação dos resultados	Análise qualitativa dos padrões identificados	Análise contextual dos resultados quantitativos	Insights sobre particularidades regionais e tendências
	Elaboração de conclusões	Síntese dos principais achados e implicações	Integração dos resultados das diferentes etapas de análise	Conclusões sobre programas prioritários e percepções por região
	Formulação de recomendações	Proposição de direcionamentos baseados na análise	Interpretação dos resultados para sugestões práticas	5 recomendações específicas para aprimoramento de políticas públicas

Observações sobre a Implementação do Pipeline

1. **Adaptações Metodológicas:** O pipeline original de Filipe Araújo foi adaptado para a análise específica da matriz PE, BA, AM, PA, mantendo os princípios fundamentais da metodologia.
2. **Limitações Identificadas:**
 - Desafio na padronização completa da terminologia entre estados
 - Variações estruturais entre as planilhas (ex: BA sem coluna de comentários)
 - Subjetividade na classificação de sentimentos em textos com linguagem técnica
3. **Resultados-Chave da Aplicação do Pipeline:**
 - Identificação de padrões regionais distintos para Norte e Nordeste
 - Mapeamento da percepção sobre programas federais por gestores locais
 - Quantificação da distribuição temática nas políticas públicas
 - Identificação de tendências de sentimento predominantemente neutras ou negativas

Temas Predominantes do Estado

Tema	PE	PA	AM	BA	Total
Programas Sociais	7	14	25	8	54
Saúde	10	27	20	14	71
Educação	15	25	28	13	81
Infraestrutura	7	21	11	9	48
Agricultura	2	14	3	1	20
Governança	30	50	51	40	171

Análise de Sentimento

Sentimento	PE	PA	AM	BA	Total
Positivo	6	6	7	7	26
Negativo	16	20	18	16	70

Sentimento	PE	PA	AM	BA	Total
Neutro	94	186	141	129	550