Decodificar não é compreender: a lacuna entre compreensão visual e interfaces de explicação de IA

Juliana Sato Yamashita
FAU-USP - DSG5028 Tópicos de Pesquisa em Design da Informação

Resumo

Interfaces de inteligência artificial explicável (XAI) podem representar corretamente os fatores que contribuíram para uma saída do modelo sem garantir que os usuários compreendam o que essa saída significa em seu contexto. Esta revisão de escopo investiga o que a literatura revela sobre como visualizações de dados e recursos explicativos apoiam a passagem da decodificação de elementos gráficos para a síntese de padrões e a compreensão contextualizada. O corpus reúne 194 artigos, recuperados por buscas sobre compreensão de visualizações, interfaces visuais de XAI e narrativa e anotação, além de uma busca manual em oito veículos de pesquisa em design. Os artigos foram atribuídos a cinco comunidades bibliográficas segundo o veículo de publicação. Uma base de 174 artigos não classificados como revisão foi analisada segundo o nível semântico priorizado por seus mecanismos de explicação, com base no modelo de Lundgard e Satyanarayan (2021). Também foi construída uma rede bipartida de citações, composta pelos artigos do corpus, 345 obras-âncora e 1.624 relações bibliográficas.

Os resultados apontam para um padrão convergente. A comunidade de IA e aprendizado de máquina apresenta a menor proporção de artigos voltados à síntese de padrões e à contextualização, com 34% dos trabalhos classificados em L3–L4. Entre os artigos dessa comunidade dedicados a interfaces de XAI, 73% concentram-se em mecanismos de decodificação. A rede de citações mostra, em paralelo, que os estudos sobre compreensão de visualizações e narrativa compartilham 113 obras-âncora, enquanto o pilar de interfaces de XAI compartilha 33 com a literatura de compreensão e 22 com a de narrativa. A conexão entre XAI e pesquisa em design é ainda menor, com 20 âncoras compartilhadas. As análises não permitem estabelecer causalidade, mas revelam uma associação consistente entre a concentração das interfaces de XAI em mecanismos de decodificação e seu afastamento bibliográfico das pesquisas dedicadas à compreensão. Conclui-se que ampliar o critério de sucesso dessas interfaces exige considerar não apenas a fidelidade da explicação, mas a compreensão que ela permite construir. Sequenciamento narrativo e ancoragem visual constituem mecanismos promissores para essa ampliação, cujos efeitos devem ser avaliados empiricamente com usuários e em contextos de domínio.

Palavras-chave: design da informação; revisão de escopo; compreensão de visualizações; inteligência artificial explicável; visualização narrativa; análise de rede de citações.

1. Introdução

Um clínico recebe um alerta indicando alto risco de deterioração do paciente. Ao lado, um gráfico SHAP mostra que determinados resultados laboratoriais, a idade e uma comorbidade contribuíram para a predição. A explicação identifica corretamente os fatores considerados pelo modelo, mas sua leitura exige compreender o sentido e a magnitude de cada contribuição. Mesmo quando o gráfico é decodificado corretamente, permanece uma questão mais importante: como relacionar esses fatores ao quadro clínico atual e decidir se o alerta exige uma intervenção imediata, acompanhamento mais próximo ou nenhuma mudança de conduta? Em contextos de saúde, essa diferença é especialmente relevante porque explicações algorítmicas podem informar condutas clínicas, priorização de atendimento e alocação de recursos, sem que a exposição dos fatores considerados pelo modelo seja suficiente para sustentar essas decisões.

A diferença entre uma explicação tecnicamente correta e uma explicação compreendida não se resolve apenas com maior rigor do modelo. Kaur et al. (2020) observaram que cientistas de dados interpretavam e utilizavam incorretamente gráficos produzidos por ferramentas comuns de interpretabilidade, apesar de sua formação técnica e familiaridade com esse tipo de sistema. As ferramentas apresentavam as visualizações previstas por seus respectivos métodos, mas isso não

garantia que os participantes as interpretassem ou utilizassem corretamente. O problema, portanto, não está apenas na modelagem. Está também no design da interface pela qual a explicação chega ao usuário.

Se a explicação é também um problema de design, o conhecimento acumulado sobre a compreensão de representações visuais deveria orientar a construção dessas interfaces. Três literaturas tratam de partes complementares desse problema. A pesquisa sobre compreensão de visualizações investiga o que permite ao leitor passar da identificação de elementos gráficos para a construção de sentido. A pesquisa em IA explicável (XAI) desenvolve interfaces destinadas a apresentar e explicar o comportamento dos modelos. Já a literatura sobre narrativa e anotação estuda mecanismos que podem orientar a leitura e relacionar elementos visuais a interpretações mais amplas.

Apesar dessa proximidade, essas literaturas se desenvolveram em grande parte separadamente. A pesquisa em XAI nem sempre incorpora o que já se sabe sobre compreensão visual, enquanto a pesquisa em design da informação e visualização narrativa raramente toma as interfaces de XAI como objeto. Esta revisão de escopo investiga esse afastamento.

A análise utiliza o modelo de quatro níveis de conteúdo semântico de Lundgard e Satyanarayan (2021). Os níveis 1 e 2 correspondem à decodificação: identificam propriedades elementares do gráfico e relações entre os elementos representados. Os níveis 3 e 4 correspondem à compreensão-para-ação: sintetizam padrões e relacionam esses padrões ao contexto de domínio.

Com base nessa distinção, esta revisão investiga como visualizações de dados e recursos explicativos apoiam a passagem entre esses níveis. A pergunta central é: o que a literatura revela sobre como essa passagem é apoiada, inclusive em interfaces de explicação de XAI?

O artigo não busca apenas identificar uma lacuna entre essas literaturas. Sua hipótese é que esse afastamento decorre da forma como a pesquisa em interfaces de explicação definiu sua própria tarefa. Grande parte dessa produção trata como resultado final aquilo que constitui apenas uma condição inicial da compreensão: mostrar, com fidelidade, o que o modelo fez. Quando a explicação termina nesse ponto, a literatura sobre compreensão deixa de parecer necessária. A rede de citações permite observar bibliograficamente essa separação.

A Seção 2 desenvolve esse argumento a partir do design da informação. A Seção 3 apresenta o método da revisão e a construção da rede de citações. A Seção 4 reúne os resultados da classificação semântica e da análise da rede. A Seção 5 discute as consequências dessa menor sobreposição bibliográfica.

2. A explicação como problema de design da informação

Antes de analisar as interfaces de explicação, é necessário definir o que constitui uma explicação bem-sucedida. Quatro posições, formuladas a partir de tradições distintas mas convergentes, oferecem esse critério: toda representação envolve interpretação; projetá-la é uma tarefa de design; seu sucesso depende da compreensão produzida no leitor; e essa compreensão é construída durante a interação, não simplesmente transmitida pela interface.

Uma visualização não apresenta dados de forma neutra. Drucker (2014) formula essa ideia por meio da distinção entre *data* e *capta* — entre aquilo que supostamente é dado e aquilo que é selecionado e construído como representação. Um gráfico de contribuição de fatores depende de decisões sobre quais variáveis mostrar, como agrupá-las, que escala utilizar e em que ordem apresentá-las. Essas decisões moldam a leitura, ainda que desapareçam na superfície final do gráfico.

A interface, portanto, não apenas exibe o resultado do modelo. Ela organiza e enquadra como esse resultado pode ser percebido. É nesse sentido que Bonsiepe (2011) situa a questão no campo do design da informação. Para o autor, o design atua na passagem entre dado, informação e conhecimento, organizando textos, imagens e diagramas em interfaces capazes de sustentar a compreensão. Bonsiepe nomeia essa operação de *mapeamento* — não o mapeamento de um

espaço físico, mas o de um espaço informacional, a disciplina de registrar, produzir e transmitir conhecimento estendida da cartografia para a interface. Uma interface de XAI não é, assim, uma camada acessória acrescentada depois que a explicação técnica estaria pronta. Suas decisões participam da própria explicação.

Frascara (2011) desloca o critério de avaliação do objeto para seus efeitos. O design da informação não deve ser julgado apenas pela correção ou aparência do artefato, mas pelo que as pessoas conseguem compreender e fazer a partir dele. Tornar uma explicação visível ou tecnicamente correta não garante que ela seja relevante para a situação do usuário. Da mesma forma, compreender separadamente os elementos de um gráfico não garante que o leitor consiga relacioná-los ao problema que precisa resolver.

A perspectiva do *sense-making* — a construção de sentido — de Dervin (1999), proposta explicitamente como teoria para o design da informação, ajuda a explicar essa diferença. A construção de sentido é um processo ativo e situado, no qual a pessoa relaciona as informações disponíveis ao que já sabe e às exigências do contexto. A compreensão não é uma quantidade pronta que passa da interface para o leitor, mas algo construído no encontro entre a pessoa, a representação e a situação.

Essa posição traz duas consequências para a análise das interfaces de explicação. A primeira é que sua avaliação não pode se limitar à fidelidade do artefato. A segunda é que decodificação e compreensão não diferem apenas em grau. Decodificar significa identificar o que os elementos representam e reconhecer relações explícitas entre eles. Compreender exige integrar essas informações, reconhecer padrões e relacioná-los ao contexto.

O modelo de Lundgard e Satyanarayan (2021) permite operacionalizar essa distinção. Os níveis 1 e 2 descrevem propriedades elementares e relações presentes na visualização e correspondem ao plano da decodificação. Os níveis 3 e 4 envolvem a síntese de padrões e sua interpretação no contexto do domínio, correspondendo à compreensão-para-ação.

Um mecanismo voltado aos níveis 1 e 2 fornece material necessário à interpretação. Um mecanismo voltado aos níveis 3 e 4 apoia sua integração e contextualização. A análise apresentada na Seção 4 examina onde cada literatura situa sua tarefa e como essa delimitação aparece nas relações bibliográficas entre XAI, compreensão de visualizações, narrativa e design da informação.

3. Método

3.1 Desenho da revisão

Esta pesquisa adota a metodologia de revisão de escopo descrita por Arksey e O'Malley (2005) e refinada por Levac, Colquhoun e O'Brien (2010). O método foi escolhido para mapear a extensão, a composição e as relações internas de um campo heterogêneo.

A revisão combina dois procedimentos. O primeiro é a leitura e classificação dos artigos incluídos, voltada a identificar o nível de compreensão apoiado pelos mecanismos de explicação e as populações estudadas. O segundo é a análise da rede de citações, empregada para examinar a base bibliográfica compartilhada pelas literaturas investigadas.

A Figura 1 apresenta o fluxo de identificação, triagem e inclusão dos registros, conforme o PRISMA-ScR (Tricco et al., 2018).

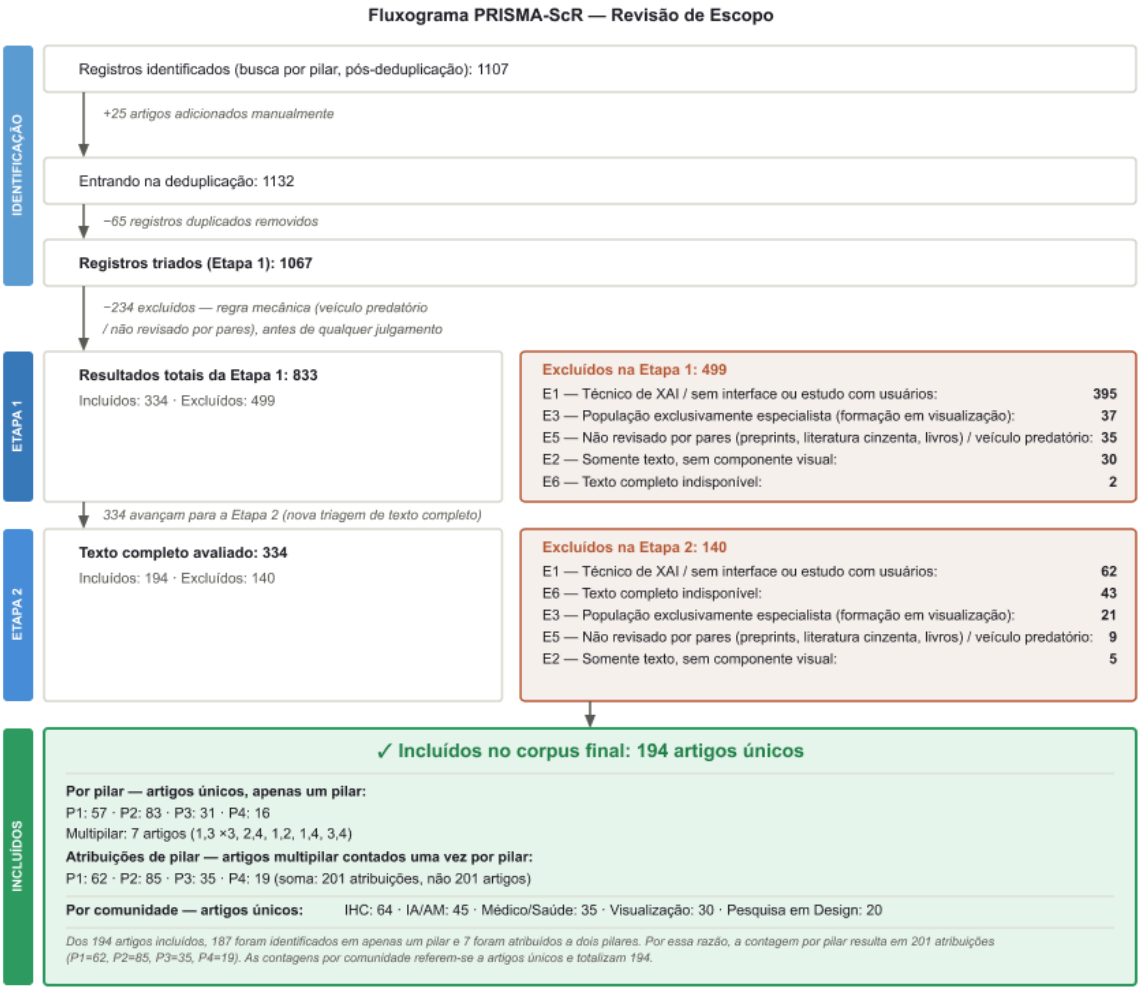


Figura 1: Fluxograma PRISMA-ScR do processo de identificação, triagem e inclusão

3.2 Estratégia de busca

A busca foi organizada em quatro pilares. O Pilar 1 (P1) reúne estudos sobre compreensão de visualizações por usuários leigos e especialistas de domínio. O Pilar 2 (P2) reúne pesquisas sobre interfaces de IA explicável — *explainable artificial intelligence* (XAI) — com componente visual. O Pilar 3 (P3) abrange narrativa e anotação como mecanismos de explicação.

O Pilar 4 (P4) foi formado por busca manual em oito veículos de pesquisa em design: *Design Studies*, *International Journal of Design*, *She Ji*, *InfoDesign*, *Visible Language*, anais do P&D Design, anais da Design Research Society (DRS) e *Estudos em Design*. Esse procedimento foi adotado porque consultas por palavras-chave no OpenAlex recuperaram essa produção de forma pouco confiável, em razão da divergência de vocabulário entre a pesquisa em design e as áreas de IHC, IA e aprendizado de máquina. A busca manual em veículos relevantes é compatível com as recomendações de Levac et al. (2010).

A identificação dos registros ocorreu no OpenAlex, escolhido também por disponibilizar as listas de referências necessárias à construção da rede. A busca do corpus foi realizada entre maio e junho de 2026, em consultas executadas em datas próximas. As listas de referências utilizadas na construção da rede de citações foram recuperadas novamente em julho de 2026, para atualizar os dados bibliográficos antes da análise.

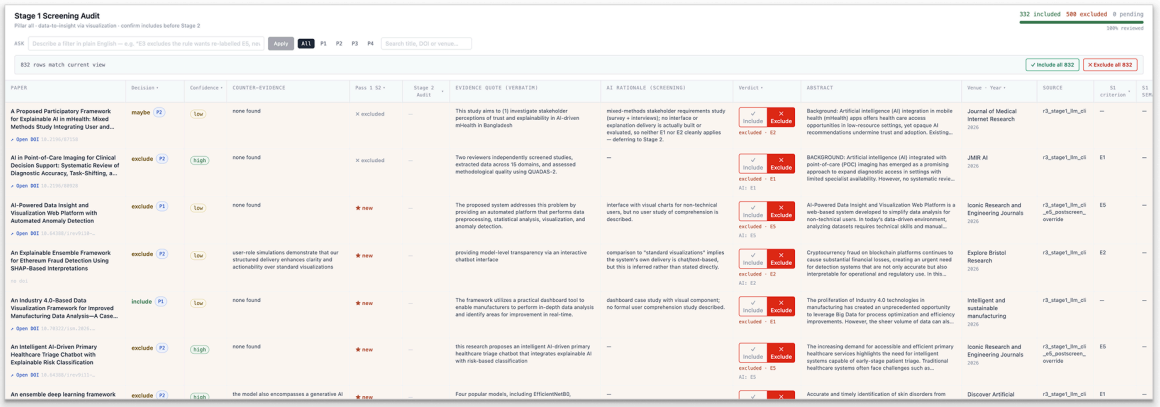
3.3 Triagem e inclusão

A triagem ocorreu em duas etapas. Na primeira, títulos e resumos foram classificados como Incluir, Excluir ou Talvez. Na segunda, os registros classificados como Incluir ou Talvez foram avaliados por texto completo.

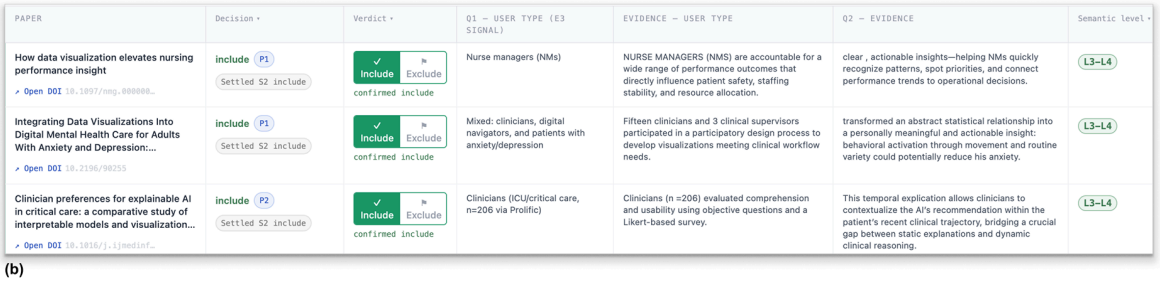
Para ser incluído, o artigo deveria tratar da compreensão, interpretação ou geração de *insight* a partir de visualizações ou saídas produzidas por IA; conter um componente visual; e ter sido publicado, em 2010 ou depois, em veículo revisado por pares.

Foram excluídos estudos dedicados exclusivamente ao desempenho técnico de modelos de XAI, sem componente de interface ou estudo com usuários; explicações apenas textuais; e trabalhos voltados exclusivamente a especialistas com treinamento comprovado em visualização de dados.

As duas etapas foram assistidas por modelos de linguagem por meio de instruções estruturadas. As decisões sugeridas foram revisadas pela autora em uma interface de auditoria desenvolvida para esse fim. Na triagem por título e resumo, a interface apresentava a decisão proposta, o nível de confiança, trechos de evidência e contraevidência, a justificativa produzida pelo modelo e os controles para confirmação ou substituição do veredito. Os campos de evidência continham citações literais extraídas do resumo, na primeira etapa, ou do texto completo do artigo, na segunda. Na triagem por texto completo, a interface apresentava também as evidências utilizadas para classificar a população de usuários, o nível semântico, o alvo de compreensão e o tipo de estudo (Figura 2). O procedimento ofereceu apoio sistemático à triagem e à extração, mas não equivale à avaliação independente por dois revisores humanos.



(a)



(b)

Figura 2. Interface de auditoria utilizada na triagem assistida por modelos de linguagem. (a) Auditoria da triagem por título e resumo, com decisão sugerida, nível de confiança, citações literais extraídas do resumo como evidência e contraevidência, justificativa do modelo e confirmação manual do veredito. (b) Auditoria por texto completo, com citações literais extraídas do artigo e utilizadas na classificação da população de usuários, do nível semântico e do alvo de compreensão.

A primeira etapa foi calibrada para favorecer a inclusão de casos incertos, que avançavam para a leitura integral. A classificação por nível semântico foi realizada apenas após o acesso ao texto completo.

A recuperação restringiu-se a artigos em acesso aberto e a cópias disponíveis em repositórios institucionais ou páginas de autores. Registros não localizados foram excluídos por indisponibilidade e contabilizados separadamente no fluxograma.

Artigos de revisão e levantamento não participaram das classificações baseadas em estudos primários, mas foram mantidos na rede de citações. Essa decisão é conservadora, pois revisões tendem a conectar diferentes vertentes de uma literatura; excluí-las poderia ampliar artificialmente a aparência de isolamento.

Após essas decisões, 174 artigos permaneceram na base de classificação por nível semântico.

3.4 Atribuição de comunidade bibliográfica

Cada artigo foi atribuído a uma de cinco comunidades bibliográficas segundo o veículo de publicação: IHC, IA/AM, Médico/Saúde, Visualização e Pesquisa em Design.

A classificação indica o circuito no qual o artigo circula e participa de uma conversa bibliográfica. Não pretende definir a formação dos autores nem uma identidade disciplinar substantiva. Suas fronteiras são, portanto, porosas: trabalhos de conteúdo semelhante podem ser classificados em comunidades diferentes conforme o veículo em que foram publicados.

3.5 Construção da rede de citações

A rede de citações foi construída para examinar a estrutura bibliográfica do corpus e as relações entre os quatro pilares. Em vez de representar apenas as citações diretas entre os artigos incluídos, a análise utiliza as obras citadas por eles como elementos de ligação. Essa escolha permite observar quais referências sustentam diferentes partes do corpus, quais são compartilhadas entre pilares e quais literaturas permanecem apoiadas em repertórios distintos.

Trata-se de uma rede bipartida, composta por dois tipos de nós. De um lado estão os 194 artigos incluídos na revisão; de outro, as obras presentes em suas listas de referências, denominadas obras-âncora. Cada aresta conecta um artigo a uma obra citada por ele. Assim, artigos que compartilham parte de sua base bibliográfica tendem a ocupar regiões próximas, enquanto aqueles apoiados em repertórios distintos tendem a se afastar.

A rede contém 345 obras-âncora e 1.624 arestas extraídas das listas de referências disponibilizadas pelo OpenAlex. Para que uma obra fosse representada como âncora, estabeleceu-se um limiar mínimo de três citações no corpus. O critério procura distinguir referências recorrentes, capazes de indicar uma base intelectual compartilhada, de referências citadas apenas uma ou duas vezes. Entre as 345 âncoras, 172 são intercruzadas, isto é, foram citadas por artigos pertencentes a dois ou mais pilares.

A visualização emprega um layout de força. Nessa configuração, os artigos são atraídos pelas obras que citam, e as obras, pelos artigos que as referenciam. Agrupamentos e distâncias resultam, portanto, dos padrões de citação, sem posicionamento manual dos nós. O objetivo não é medir a influência individual de cada artigo, mas tornar visível o grau de terreno bibliográfico comum entre as literaturas investigadas.

A codificação visual utiliza dois canais fixos. A cor representa o pilar em que cada artigo foi identificado, e o tamanho das âncoras corresponde ao número de citações recebidas no corpus. Outras classificações — nível semântico, comunidade bibliográfica e tipo de âncora — são apresentadas por interação, sem alterar a posição dos nós.

Essa decisão evita sobrepor vários códigos visuais em uma rede já densa. Como mostra Ware (2020), a busca por uma conjunção de características tende a exigir processamento serial, mais lento do que a identificação de uma única característica visual. Ao selecionar uma categoria, os

nós correspondentes são destacados; no caso das âncoras compartilhadas entre pilares, o destaque é feito por um anel amarelo de contorno.

A interface também permite destacar as âncoras intercruzadas, isto é, citadas por artigos pertencentes a dois ou mais pilares. Quando esse destaque é ativado, essas obras recebem um anel de contorno, enquanto os demais nós e arestas permanecem visíveis. A visão geral da rede e a visualização com as âncoras intercruzadas destacadas são apresentadas nas Figuras 3 e 4.

A rede representa referências compartilhadas, e não equivalência de conteúdo. Dois artigos podem citar a mesma obra por razões distintas, assim como trabalhos próximos em conteúdo podem apoiar-se em tradições bibliográficas diferentes. Os limites dessa inferência são retomados na Seção 6.

3.6 Classificação por nível semântico e população de usuários

A leitura integral produziu duas classificações: o nível semântico visado pelo mecanismo de explicação e a população de usuários estudada.

A classificação semântica baseia-se no modelo de Lundgard e Satyanarayan (2021). Ela distingue mecanismos voltados principalmente à decodificação, correspondentes aos níveis 1 e 2, daqueles destinados à síntese de padrões ou à interpretação contextualizada, correspondentes aos níveis 3 e 4. As 20 revisões foram excluídas dessa análise, resultando em uma base de 174 artigos.

A classificação por população foi aplicada aos 146 artigos empíricos ou de métodos mistos com participantes. Foram excluídos dessa análise 20 artigos de revisão, 25 artigos de pesquisa em design e três trabalhos teóricos. Essas exclusões não se aplicam à classificação semântica, pois um artigo pode propor um mecanismo de explicação mesmo sem avaliá-lo com participantes.

As populações foram verificadas no texto completo e classificadas segundo regras explícitas, com registro da evidência utilizada.

4. Resultados

4.1 Visão geral do corpus

O processo de busca e triagem resultou em 194 artigos, distribuídos entre os três pilares recuperados por consultas no OpenAlex e o quarto pilar, formado pela busca manual em veículos de pesquisa em design. As publicações abrangem o período de 2012 a 2026, com concentração nos anos mais recentes: 143 artigos, ou 74% do corpus, foram publicados a partir de 2022. O maior número de publicações ocorreu em 2025, com 47 artigos, seguido por 2024, com 43.

Quanto ao tipo de estudo, o corpus contém 108 artigos empíricos, 38 de métodos mistos, 25 de pesquisa em design, 20 revisões e três trabalhos teóricos. As análises seguintes utilizam os subconjuntos definidos na Seção 3.6.

Tabela 1. Distribuição dos artigos por comunidade bibliográfica e pilar (194 artigos únicos; 201 atribuições de pilar)

Comunidade	P1	P2	P3	P4	Total
IHC	18	30	17	1	64
IA/AM	5	36	4	0	45
Médico/Saúde	18	17	0	0	35
Visualização	18	1	13	0	30
Pesquisa em Design	3	1	1	18	20
Total	62	85	35	19	194

Nota: sete artigos foram atribuídos a dois ou mais pilares e, por isso, aparecem em mais de uma coluna. As células de P1–P4 somam 201 atribuições, embora o corpus contenha 194 artigos únicos. Os totais por comunidade contam cada artigo uma única vez.

Destacam-se a ausência de artigos da comunidade Médico/Saúde no P3 e a presença de apenas um artigo de Pesquisa em Design no P2. Essas ausências são retomadas nas análises seguintes.

4.2 A divisão por população de usuários

Os pilares diferem também pelas populações estudadas. P3 e P4 concentram-se em audiências leigas: nenhum de seus estudos empíricos investigou exclusivamente especialistas de domínio, embora três trabalhos tenham incluído esse perfil em amostras mistas. P1 e P2, em contraste, incluem especialistas de domínio com maior frequência.

A comunidade Médico/Saúde apresenta a maior presença desse público: 15 dos 22 artigos empíricos estudaram especialistas de domínio. Contudo, essa comunidade não possui nenhum trabalho no P3. Assim, a comunidade que mais investiga usuários com conhecimento especializado sobre o problema permanece ausente da literatura sobre narrativa e anotação.

A população estudada também se associa ao nível semântico do conteúdo. Entre os artigos voltados ao público geral, 27% visam conteúdo L3–L4. Entre os trabalhos com especialistas de domínio, essa proporção chega a 49%. Essa diferença, porém, não explica o contraste entre os pilares. Em todas as categorias de população — usuários leigos, público geral, especialistas de domínio e amostras mistas —, o P1 apresenta maior proporção de conteúdo L3–L4 do que o P2.

Mecanismos de comunicação e populações especialistas permanecem, portanto, parcialmente separados. A literatura que desenvolve narrativa e anotação raramente avalia esses mecanismos com especialistas de domínio, enquanto os estudos voltados a esse público os incorporam com pouca frequência.

4.3 A concentração em decodificação

Os artigos foram classificados segundo o nível semântico priorizado por seus mecanismos de explicação, com base no modelo de Lundgard e Satyanarayan (2021). As 20 revisões foram excluídas dessa análise por não relatarem mecanismos ou estudos primários próprios.

Dos 174 artigos classificados, 89 concentram-se na decodificação em L1–L2, 81 visam à compreensão-para-ação em L3–L4 e quatro combinam os dois planos. A distribuição agregada é quase equilibrada, mas encobre diferenças entre as comunidades.

Tabela 2. Distribuição dos artigos por nível semântico e comunidade bibliográfica

Comunidade	L1–L2	L3–L4	Misto	Total	% L3–L4
Médico/Saúde	12	16	0	28	57%
Visualização	11	14	2	27	56%
IHC	31	29	2	62	48%
Pesquisa em Design	10	9	0	19	47%
IA/AM	25	13	0	38	34%
Total	89	81	4	174	48%

Nota: quatro artigos, dois de IHC e dois de Visualização, combinam os dois planos semânticos e compõem uma categoria própria. As porcentagens de L3–L4 foram calculadas sobre a soma de L1–L2 e L3–L4, com exclusão dos artigos mistos

IA/AM apresenta a menor proporção de artigos voltados a L3–L4: 34%, diante de 47% a 57% nas demais comunidades. Dentro do P2, a concentração em decodificação é ainda maior. Entre os 30 artigos de IA/AM sobre interfaces de XAI, 22, ou 73%, priorizam mecanismos L1–L2; apenas oito visam à síntese de padrões ou à contextualização em L3–L4.

Esse resultado não indica necessariamente que esses trabalhos prometam compreensão contextualizada e deixem de oferecê-la. Em muitos casos, a tarefa é definida de forma mais restrita: tornar visíveis os fatores, relações ou regiões associados à saída do modelo. Quando a explicação termina nesse ponto, a síntese desses elementos e sua relação com o contexto do usuário permanecem fora de seu escopo.

Essa diferença aparece também em estudos concretos do corpus. Okolo et al. (2024) observaram que trabalhadoras comunitárias de saúde em áreas rurais da Índia tiveram dificuldade para interpretar versões simplificadas de LIME e SHAP: as participantes confundiram importância de atributos com sintomas da doença e interpretaram elementos visuais segundo associações prévias de cor e gravidade. Em outro estudo com trabalhadoras comunitárias de saúde no mesmo contexto regional, Solano-Kamaiko et al. (2024) empregaram explicações exploráveis que permitiam modificar medidas, comparar casos e observar mudanças na previsão. Os autores relatam que essas interações apoiaram modelos mentais mais nuançados sobre o funcionamento da IA e ampliaram a capacidade das participantes de contestar suas decisões. O contraste ilustra a diferença entre disponibilizar os componentes de uma explicação e projetar recursos que apoiem sua integração e seu uso.

A seção seguinte examina como essa delimitação aparece na base bibliográfica dos artigos.

4.4 A menor sobreposição bibliográfica do P2

A rede de citações mostra diferenças na sobreposição bibliográfica entre os pilares. Na Figura 3, os artigos do P2, representados em verde e concentrados sobretudo na região direita da rede, formam um agrupamento relativamente coeso. Os artigos de P1, P3 e P4, representados respectivamente em azul, laranja e roxo, ocupam regiões mais próximas e parcialmente sobrepostas à esquerda e ao centro.

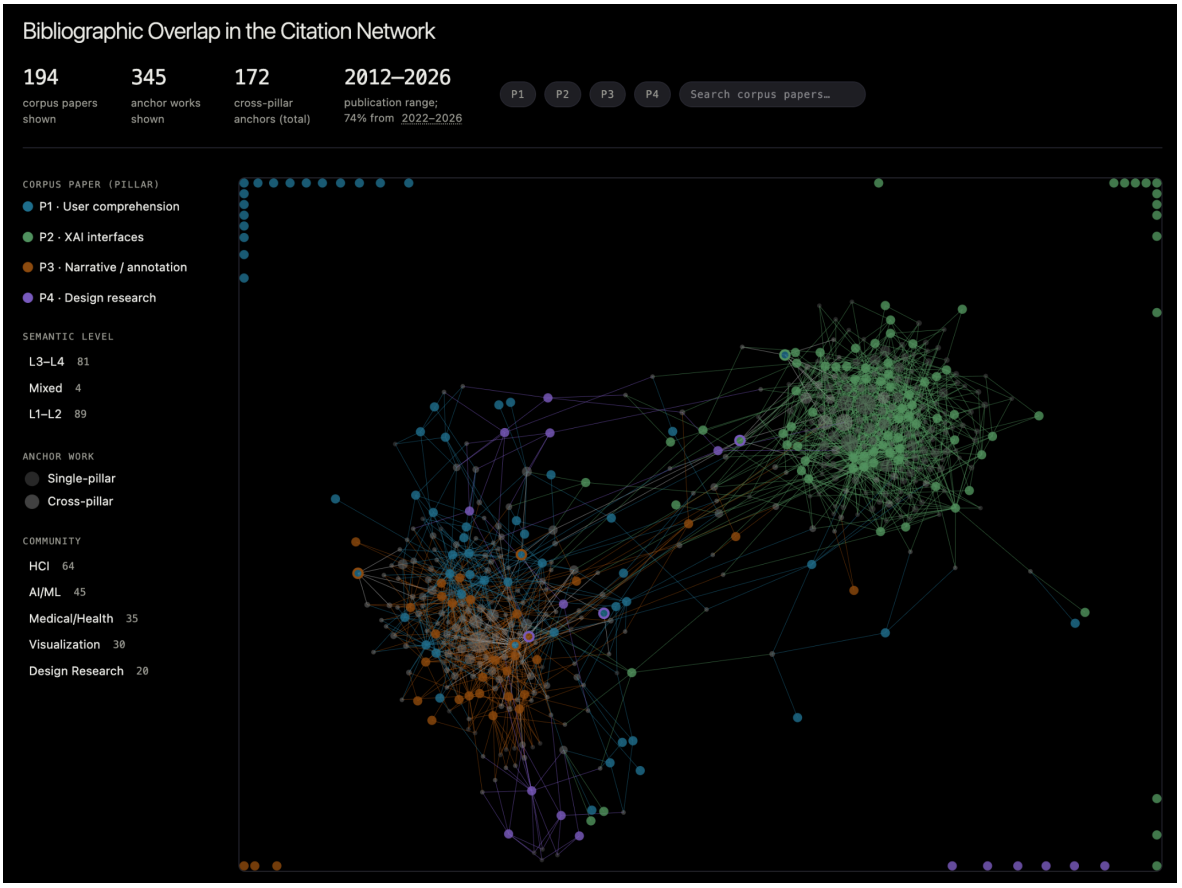


Figura 3. Visão geral da rede bipartida de citações. Os artigos do corpus são coloridos segundo o pilar em que foram identificados: P1 em azul, P2 em verde, P3 em laranja e P4 em roxo. As obras-âncora aparecem em cinza, com tamanho proporcional ao número de citações recebidas no corpus. A posição dos nós resulta de um layout de força baseado nas relações de citação.

Como o posicionamento dos nós resulta das referências compartilhadas, essa configuração indica que os artigos de P2 se apoiam em um repertório bibliográfico relativamente coeso, mas menos conectado às referências mobilizadas pelos outros pilares. A separação espacial, contudo, não constitui por si só uma medida dessa diferença; as contagens de âncoras compartilhadas apresentadas a seguir fornecem a evidência quantitativa.

Das 345 obras-âncora representadas na rede, 172 são intercruzadas, isto é, foram citadas por artigos pertencentes a dois ou mais pilares. A Tabela 3 apresenta as conexões por par.

Tabela 3. Obras-âncora compartilhadas por par de pilares

Ponte	Literaturas conectadas	Âncoras (n)
P1 + P3	Compreensão de visualizações ↔ Narrativa/anotação	113
P1 + P2	Compreensão de visualizações ↔ Interfaces de XAI	33
P2 + P3	Interfaces de XAI ↔ Narrativa/anotação	22
P1 + P4	Compreensão de visualizações ↔ Pesquisa em design	32
P3 + P4	Narrativa/anotação ↔ Pesquisa em design	28
P2 + P4	Interfaces de XAI ↔ Pesquisa em design	20

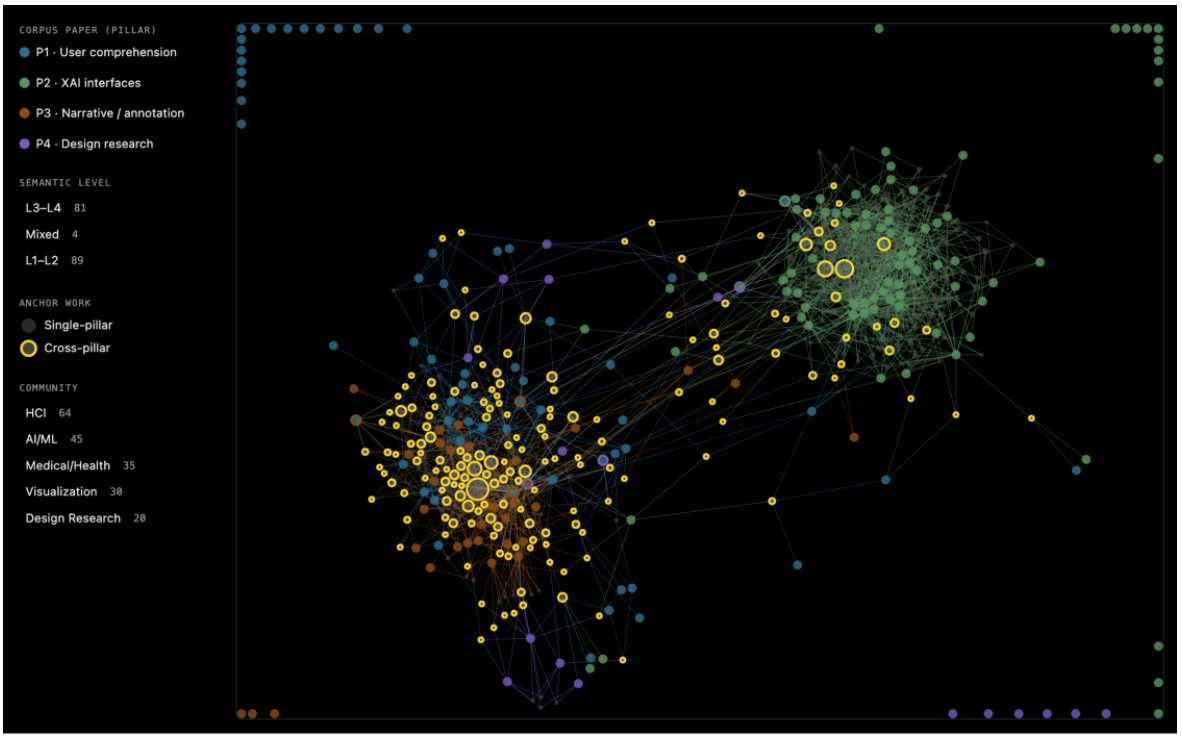
Nota: uma mesma âncora pode conectar mais de dois pilares e, portanto, ser contada em mais de um par. Por essa razão, as contagens da tabela não totalizam 172.

A principal conexão ocorre entre P1 e P3, com 113 âncoras compartilhadas. P2 compartilha apenas 33 âncoras com P1 e 22 com P3. A menor conexão é entre P2 e P4, com 20 obras compartilhadas, embora ambos tratem da apresentação e interpretação de conteúdos visuais.

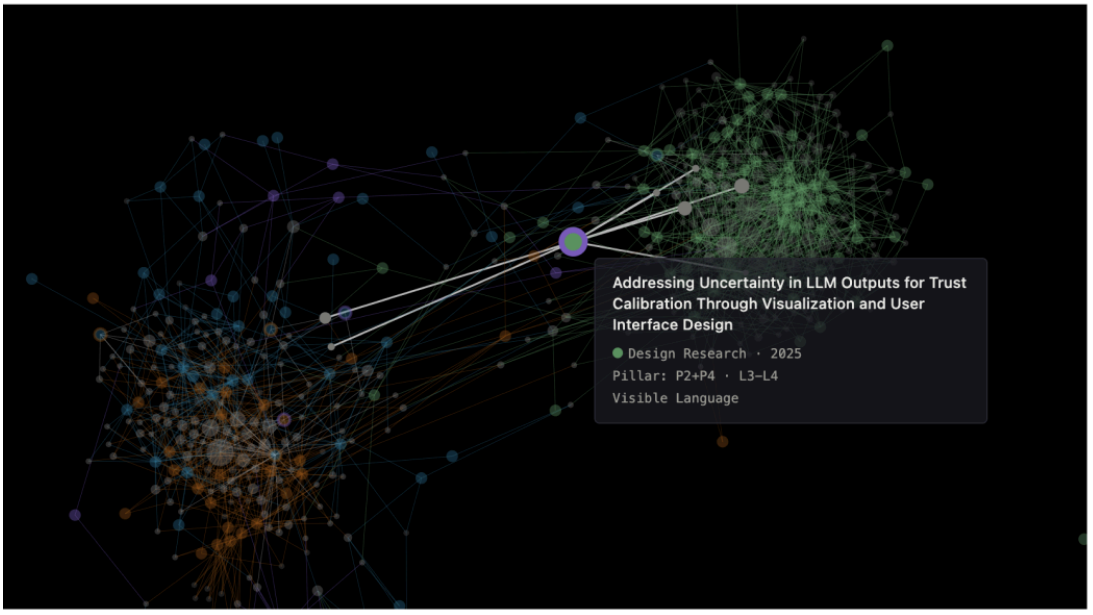
A menor sobreposição entre P2 e P4 não significa ausência de diálogo. Armstrong et al. (2025), publicado em um veículo de pesquisa em design e atribuído aos dois pilares, constitui uma ponte explícita entre essas literaturas. Na rede, o artigo ocupa uma posição intermediária e se conecta a obras-âncora compartilhadas entre a região do P2 e os demais pilares (Figura 4b). O trabalho combina literatura de XAI, confiança em automação, visualização de incerteza e design de interfaces para propor convenções visuais, um framework de incerteza e protótipos destinados a apoiar a validação de resumos produzidos por LLMs. O caso mostra que a aproximação entre design e XAI já ocorre em trabalhos pontuais, embora ainda não constitua uma base bibliográfica amplamente compartilhada.

O padrão não decorre apenas do tamanho dos conjuntos. P2 possui 202 âncoras, o maior conjunto da rede, enquanto P1 e P3 possuem 149 cada um e P4 possui 52. Apesar de menor, P4 compartilha 32 âncoras com P1 e 28 com P3, diante de apenas 20 com P2.

A normalização confirma a diferença. P1 compartilha com P3 113 das 149 âncoras deste último, ou 75,8%. P2 compartilha com P3 apenas 22, equivalentes a 14,8%. Considerando o conjunto de P1, 75,8% de suas âncoras são compartilhadas com P3, diante de 22,1% com P2.



(a)



(b)

Figura 4. Âncoras compartilhadas entre pilares e exemplo de ponte bibliográfica entre P2 e P4. (a) Rede completa com destaque, em amarelo, das obras-âncora citadas por artigos pertencentes a dois ou mais pilares. (b) Detalhe da posição de Armstrong et al. (2025), artigo atribuído aos pilares P2 e P4, com destaque das obras-âncora às quais se conecta. O caso exemplifica uma aproximação pontual entre interfaces de XAI e pesquisa em design.

Na visualização com as âncoras intercruzadas destacadas (Figura 4a), observa-se maior concentração dessas obras na região ocupada por P1, P3 e P4, enquanto a região do P2 apresenta menos pontes com o restante da rede.

Os resultados mostram, assim, três diferenças relacionadas: as literaturas estudam populações distintas, concentram-se em níveis semânticos diferentes e apresentam graus desiguais de sobreposição bibliográfica. O P2 mantém conexões com os demais pilares, mas em menor proporção do que a observada entre P1, P3 e P4.

5. O que a lacuna custa

Os resultados apontam para o mesmo limite em dois planos. No plano semântico, grande parte da literatura de XAI concentra-se em tornar visíveis os fatores, relações ou regiões associados à saída do modelo. No plano bibliográfico, essa produção compartilha menos referências com os estudos sobre compreensão, narrativa e design da informação.

As duas linhas não demonstram uma relação causal, mas são coerentes entre si. Quando a explicação é definida como apresentação fiel dos componentes de uma saída, a síntese desses componentes e sua contextualização permanecem fora da tarefa principal. Nesse enquadramento, a literatura sobre compreensão deixa de ser instrumentalmente necessária para que o objetivo declarado seja alcançado.

A menor sobreposição bibliográfica registrada pela rede não precisa, portanto, ser interpretada como negligência. Ela decorre do limite estabelecido para o próprio objeto. Os elementos da explicação são disponibilizados, mas o trabalho de integrá-los e relacioná-los à situação continua a cargo do usuário. Em contextos como a saúde e a decisão pública, essa compreensão importa porque explicações algorítmicas podem participar de avaliações e escolhas com consequências concretas, mesmo quando não determinam diretamente a decisão.

A participação de especialistas de domínio no processo de projeto é importante, mas não resolve, por si só, os problemas de compreensão. Mesmo quando usuários ajudam a definir o conteúdo e o contexto de uso, permanece necessário projetar como a visualização e seus recursos explicativos apoiam a passagem da identificação dos elementos gráficos para sua interpretação contextualizada.

Uma revisão sobre geração de linguagem natural para visualizações identificou um problema relacionado: a maioria dos sistemas apresenta o texto separadamente do gráfico, enquanto a vinculação automática entre afirmações e elementos visuais permanece rara (Hoque & Islam, 2025). A separação bibliográfica encontra, assim, um paralelo nas próprias interfaces construídas.

O afastamento também ocorre no sentido inverso. A presença de apenas um artigo de Pesquisa em Design no P2 indica que interfaces de XAI ainda são pouco tomadas como objeto por uma literatura dedicada à comunicação e à interpretação de informações visuais.

Parte dessa distância é expressa pelo vocabulário. Enquanto XAI mobiliza termos como fidelidade, importância de atributos, saliência e interpretabilidade, o design aborda compreensão, construção de sentido, enquadramento e relação entre informação e leitor. A divergência terminológica dificulta tanto a recuperação bibliográfica quanto o reconhecimento de problemas compartilhados.

Os resultados delimitam, portanto, um espaço de design ainda pouco explorado. A questão não é corrigir mecanismos que já pretendam apoiar L3–L4, mas investigar como recursos explicativos associados à visualização podem apoiar a passagem da exposição dos componentes para sua síntese e contextualização.

A literatura examinada aponta dois mecanismos para essa investigação. O primeiro é o sequenciamento do conteúdo por nível semântico, organizando a explicação para avançar de propriedades elementares a relações, padrões e contexto. Hullman e Diakopoulos (2011) mostram que decisões de sequência e enquadramento influenciam a interpretação de visualizações narrativas.

O segundo é a ancoragem visual, que vincula cada afirmação ao elemento gráfico correspondente. Essa operação aproxima linguagem e representação no mesmo espaço de leitura e pode reduzir o custo de integração produzido pela busca alternada entre uma afirmação e seu referente visual (Franconeri et al., 2021).

Esses mecanismos não constituem soluções garantidas. Sequenciar também enquadra o que será percebido primeiro, e ancorar uma afirmação pode reduzir a saliência dos elementos não

destacados. Sua eficácia deve, portanto, ser avaliada empiricamente, em contextos de domínio e com medidas capazes de distinguir decodificação de compreensão contextualizada.

6. Limitações

A atribuição das comunidades bibliográficas pelo veículo de publicação permite identificar circuitos de circulação, mas produz fronteiras porosas entre áreas.

A rede de citações indica terreno bibliográfico compartilhado, não convergência conceitual. Uma mesma obra pode ser mobilizada de formas distintas, e literaturas próximas podem apoiar-se em repertórios diferentes. A análise mostra, portanto, associações bibliográficas, não uma relação causal entre a menor sobreposição bibliográfica do Pilar 2 e sua concentração em mecanismos de decodificação.

O corpus depende da cobertura do OpenAlex e, nas análises baseadas em leitura integral, da disponibilidade dos textos completos. Registros ausentes ou não recuperados podem alterar sua composição e a distribuição dos níveis semânticos. A rede também permanece condicionada à completude das listas de referências fornecidas pela base. Embora a cobertura de referências do OpenAlex seja comparável à do Web of Science e do Scopus para publicações recentes presentes nas três bases, sua cobertura e seus metadados variam entre áreas e versões da base (Culbert et al., 2025).

Os procedimentos de recuperação não foram equivalentes: P1, P2 e P3 foram formados por consultas no OpenAlex, enquanto P4 resultou de busca manual em veículos selecionados.

A triagem foi assistida por modelos de linguagem e auditada por uma única revisora humana, o que não equivale à avaliação independente por dois revisores.

A classificação semântica registra o nível visado pelo mecanismo segundo a descrição do artigo, e não a compreensão efetivamente alcançada pelos usuários.

Os resultados devem ser interpretados como padrões observados no corpus analisado, não como uma descrição exaustiva ou causal das literaturas investigadas.

7. Conclusão

Esta revisão de escopo examinou 194 artigos distribuídos entre cinco comunidades bibliográficas e quatro pilares temáticos. As análises semântica e bibliográfica convergem: as interfaces de XAI concentram-se mais em mecanismos de decodificação e apoiam-se menos nas literaturas sobre compreensão, narrativa e design da informação.

Grande parte dessa produção define a explicação como apresentação dos fatores, relações ou regiões associados à saída do modelo. Esse objetivo é coerente com os mecanismos desenvolvidos, mas deixa fora da tarefa a síntese dos elementos e sua relação com o contexto do usuário.

A contribuição do design da informação está em ampliar o critério de sucesso. Uma explicação não deve ser avaliada apenas pela correção do que apresenta, mas também pela compreensão que permite construir. Decodificar fornece elementos necessários à interpretação; não garante que o leitor consiga integrá-los, reconhecer um padrão ou relacioná-lo à situação em que precisa agir.

Sequenciamento narrativo e ancoragem visual constituem mecanismos promissores para investigar essa passagem, mas seus efeitos não devem ser presumidos. Precisam ser avaliados com usuários e em contextos de domínio.

Enfrentar a lacuna documentada exige, portanto, mais do que ampliar a circulação de referências entre comunidades. Exige rever o que se considera uma explicação bem-sucedida. Quando a compreensão do usuário integra esse critério, o conhecimento produzido pela pesquisa

em visualização, narrativa e design da informação deixa de ser periférico e se torna parte do projeto de interfaces de explicação.

8. Declaração sobre o uso de inteligência artificial

Ferramentas de inteligência artificial generativa foram utilizadas em diferentes etapas deste trabalho. Durante a revisão, modelos de linguagem auxiliaram a triagem de títulos, resumos e textos completos a partir de critérios e instruções estruturados. Todas as decisões foram auditadas pela autora e, quando necessário, corrigidas manualmente. Durante a redação, o ChatGPT (OpenAI) e o Claude (Anthropic) foram empregados para revisar clareza, concisão e organização e para reformular trechos produzidos pela autora. O Claude Code (Anthropic) foi utilizado como apoio à programação e à prototipagem da visualização da rede de citações. As decisões metodológicas, classificações, análises, escolhas de design, interpretações e conclusões permaneceram sob responsabilidade da autora, que também revisou o texto final, a visualização e as referências citadas.

Referências

ARKSEY, Hilary; O'MALLEY, Lisa. Scoping studies: towards a methodological framework. International Journal of Social Research Methodology, v. 8, n. 1, p. 19–32, 2005.

ARMSTRONG, Helen; ANDERSON, Ashley L.; PLANCHART, Rebecca; BAIDOO, Kweku; PETERSON, Matthew. Addressing uncertainty in LLM outputs for trust calibration through visualization and user interface design. Visible Language, v. 59, n. 2, p. 176–217, 2025.

BONSIEPE, Gui. Design, cultura e sociedade. São Paulo: Blucher, 2011.

CULBERT, Jack et al. Reference coverage analysis of OpenAlex compared to Web of Science and Scopus. Scientometrics, v. 130, p. 2475–2492, 2025. DOI 10.1007/s11192-025-05293-3.

DERVIN, Brenda. Chaos, order, and sense-making: a proposed theory for information design. In: JACOBSON, Robert (Ed.). Information design. Cambridge, MA: MIT Press, 1999. p. 35–57.

DRUCKER, Johanna. Graphesis: visual forms of knowledge production. Cambridge, MA: Harvard University Press, 2014.

FRASCARA, Jorge. ¿Qué es el diseño de información? Buenos Aires: Ediciones Infinito, 2011.

FRANCONERI, Steven L. et al. The science of visual data communication: what works. Psychological Science in the Public Interest, v. 22, n. 3, p. 110–161, 2021.

HOQUE, Enamul; ISLAM, Mohammed Saidul. Natural language generation for visualizations: state of the art, challenges and future directions. Computer Graphics Forum, v. 44, n. 1, e15266, 2025. DOI 10.1111/cgf.15266.

HULLMAN, Jessica; DIAKOPOULOS, Nicholas. Visualization rhetoric: framing effects in narrative visualization. IEEE Transactions on Visualization and Computer Graphics, v. 17, n. 12, p. 2231–2240, 2011.

KAUR, Harmanpreet et al. Interpreting interpretability: understanding data scientists' use of interpretability tools for machine learning. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. New York: ACM, 2020. p. 1–14.

LEVAC, Danielle; COLQUHOUN, Heather; O'BRIEN, Kelly K. Scoping studies: advancing the methodology. Implementation Science, v. 5, art. 69, 2010.

LUNDGARD, Alan; SATYANARAYAN, Arvind. Accessible visualization via natural language descriptions: a four-level model of semantic content. IEEE Transactions on Visualization and Computer Graphics, v. 28, n. 1, p. 1073–1083, 2021.

OKOLO, Chinasa T.; AGARWAL, Dhruv; DELL, Nicola; VASHISTHA, Aditya. “If it is easy to understand, then it will have value”: examining perceptions of explainable AI with community health workers in rural India. Proceedings of the ACM on Human-Computer Interaction, v. 8, CSCW1, art. 71, 2024. DOI 10.1145/3637348.

SOLANO-KAMAIKO, Ian René; MISHRA, Dibyendu; DELL, Nicola; VASHISTHA, Aditya. Explorable explainable AI: improving AI understanding for community health workers in India. In: Proceedings of the CHI Conference on Human Factors in Computing Systems. New York: ACM, 2024. DOI 10.1145/3613904.3642733.

TRICCO, Andrea C. et al. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of Internal Medicine, v. 169, n. 7, p. 467–473, 2018. DOI 10.7326/M18-0850.

WARE, Colin. Information visualization: perception for design. 4. ed. Burlington: Morgan Kaufmann, 2020.