information

MDPI

Review

Entity Resolution Using Transformer-Based Language Models: A Systematic Scoping Review

Mohammad Beheshti [1,2,3], Maryam Seifaddini [1,2,3], Amir Erfan Zareei Shams Abadi [2], Karan Karthik [1,2], Tarun Mummidi Ramesh Kumar [1,2], Suzanne Austin Boren [2,3] and Iris Zachary [1,2,3,*]

1 Missouri Cancer Registry and Research Center, University of Missouri, Columbia, MO 65211, USA; mbwnh@missouri.edu (M.B.); mshb6@missouri.edu (M.S.); karankarthik@missouri.edu (K.K.); tmgbw@missouri.edu (T.M.R.K.)
2 MU Institute for Data Science and Informatics, University of Missouri, Columbia, MO 65211, USA; azzgg@missouri.edu (A.E.Z.S.A.); borens@missouri.edu (S.A.B.)
3 College of Health Sciences, University of Missouri, Columbia, MO 65211, USA
* Correspondence: zacharyi@missouri.edu

Abstract

Entity resolution (ER) is fundamental to integrating heterogeneous data, which traditional approaches address through deterministic rule-based methods and probabilistic record linkage. We conducted a systematic scoping review following PRISMA-ScR to characterize the use of transformer-based language models for ER. Five databases were searched, and 155 studies were included for synthesis. The literature expanded sharply after 2023, with 55% of included studies published in 2025 or 2026. General-purpose matching was the most common entity focus, followed by product/e-commerce. Healthcare applications were especially scarce, with only one study applying ER to the patient/healthcare domain. Among studies that performed blocking, embedding-based similarity was most common, followed by string-based approaches. Classification-head approaches remained the most common matching approach, followed by prompt-based approaches. Encoder-only models remained the most widely evaluated architecture, while decoder-only models grew increasingly prominent from 2023 onward. Full-parameter fine-tuning was the predominant learning strategy, followed by zero-shot and few-shot prompting. Among studies classified as general-purpose, nearly one-third were evaluated on only one or two entity types, limiting their generalizability. Reported best F1 scores varied across benchmarks, with no consistent advantage for encoder-only versus decoder-only architectures. These findings support the need for more diverse, end-to-end evaluations that consider efficiency and robustness alongside accuracy.

Keywords: entity resolution; record linkage; deduplication; language models; generative AI; artificial intelligence

Academic Editor: Sven Groppe

Received: 23 July 2026
Revised: 10 September 2026
Accepted: 15 September 2026
Published: 18 September 2026

1. Introduction

Entity resolution (ER), also known as entity matching (EM), record linkage, or deduplication, is a fundamental task in data management that involves identifying and matching records that refer to the same real-world entities across heterogeneous datasets [1]. It underlies applications ranging from linking patient records across health systems to deduplicating product listings in e-commerce catalogs. In healthcare specifically, accurate patient matching underpins safe care coordination and reliable public health surveillance, and unresolved duplicate or fragmented patient records remain a persistent, costly problem across

health systems [2,3]. Yet ER remains difficult in practice, since records can differ in structure, terminology, completeness, and quality, and large-scale applications often involve substantial data volumes [4,5].

Traditional approaches to ER fall broadly into two families. Deterministic (rule-based) methods declare a match when a pair of records satisfies a fixed set of exact- or threshold-based field-comparison rules, whereas probabilistic approaches, following the classical Fellegi–Sunter framework, instead estimate the likelihood that two records refer to the same entity from the statistical pattern of agreement and disagreement across their fields [6,7]. Deterministic methods depend on manually authored, exhaustive rule sets that become increasingly difficult to maintain as the number of edge cases grows, while the classical machine learning methods often used alongside them instead depend on extensive, manually engineered features [8]. Both families of approaches rely heavily on domain-specific knowledge, which limits their ability to handle unstructured data, heterogeneous sources, and shifting data patterns as ER applications grow larger and more diverse. This gap has pushed the field toward approaches that learn richer, more flexible representations of records directly from data, with less reliance on manually engineered features and rules.

This shift has been driven primarily by the transformer architecture, which uses a self-attention mechanism to relate every token in a sequence to every other token, in place of the strictly sequential processing used by earlier recurrent models, enabling substantially more effective learning of contextual relationships within text [9]. Combined with large-scale self-supervised pretraining on unlabeled text corpora, this architecture allows a single model to learn general-purpose language representations that can then be adapted to a wide range of downstream tasks [10]. Throughout this review, we use *language models* as a general term covering all such transformer-based models and reserve *large language models (LLMs)* specifically for generative models that are typically adapted through prompting.

Language models have demonstrated the ability to capture contextual and semantic relationships within text [11–13], which are directly relevant to ER, where matching decisions often depend on similarities and relationships that may not be captured by surface-level features alone. More broadly, interest in applying these capabilities across science has only accelerated since the public release of ChatGPT (based on GPT-3.5) in late 2022, which, along with the generative models that followed, has driven a rapid, well-documented surge in research applying LLMs across scientific fields [14]. ER has been no exception, with a growing body of research applying transformer-based language models across a range of datasets, domains, and methodological settings [15–17].

Prior reviews have tracked the development of neural and language-model-based approaches to entity matching. Barlaug and Gulla [18] provided an early and influential survey of neural approaches to entity matching, but their review predates the rapid emergence of generative and prompt-based language models. More recently, Awick and Marx Gómez [19] conducted a structured literature review of entity matching using language models, covering 54 studies through late 2024. Two gaps remain, though. First, the literature has continued to expand rapidly, particularly with the growing use of generative models and the increasing application of language models in specialized domains. Second, neither review gives sustained attention to biomedical and clinical applications, or synthesizes reported performance in a manner that accounts for differences among datasets and benchmarks. These developments motivate an updated synthesis of the expanding literature on transformer-based language models for ER.

This scoping review follows the PRISMA Extension for Scoping Reviews (PRISMA-ScR) guidelines to identify and pool research on transformer-based language models applied to entity resolution. We aimed to: (1) examine the models, methodologies, and datasets used across studies, including the domains and application areas represented,

with particular attention to clinical and biomedical applications; (2) compare reported performance metrics across studies using common benchmarks; and (3) identify emerging trends and research gaps in the field. Together, these analyses give an updated picture of how transformer-based language models have been applied to entity resolution across models, architectures, datasets, methods, and application domains, and point to where further research is needed.

2. Methods

This scoping review was conducted following PRISMA-ScR to ensure transparency and rigor throughout the review process [20]. Given the pace of growth in this literature and the heterogeneity of benchmarks currently in use, this review is intended to map the breadth of the literature, including the range of models, methods, benchmarks, and application domains represented, making it possible to track shifts in architecture and methodology and to identify concentration and gaps in evaluation coverage across a large and fast-changing body of work. Because evaluation protocols and reporting conventions vary widely across the included studies, this review takes a descriptive approach throughout, consistent with the aims of a scoping review.

A review protocol specifying the eligibility criteria, search strategy, and data extraction procedures was developed prior to conducting this review but was not registered on a public registry.

2.1. Search Strategy

The search was conducted across five academic databases: PubMed, Scopus, Web of Science, ACM Digital Library, and IEEE Xplore. Each database was searched on 17 August 2026. A comprehensive search strategy was developed around two concept blocks, combined with the Boolean operator AND:

> (“large language model*” OR LLM OR “language model*” OR “transformer model*” OR transformer OR “generative AI” OR “foundation model*” OR “small language model*” OR BERT OR RoBERTa) AND (“entity resolution” OR “entity matching” OR “identity resolution” OR “record linkage” OR “data linkage” OR “probabilistic linkage” OR “data deduplication” OR “record deduplication” OR deduplication OR “duplicate detection” OR “patient matching” OR “person matching” OR “master patient index” OR “patient deduplication”)

The exact query syntax, including field restrictions, wildcard conventions, and phrase-matching behavior, was adapted to each database’s search interface; the full search strings used for each database are provided in Appendix A. Searches were restricted to studies published from 2017 onward, corresponding to the introduction of the transformer architecture [9], and to English-language publications. Searches were limited to peer-reviewed journal articles, conference papers, and reviews; preprints (e.g., arXiv, bioRxiv, medRxiv) were excluded from all databases.

2.2. Study Selection

Four reviewers, working in two pairs, independently screened the titles and abstracts of the studies against the predefined inclusion and exclusion criteria. The screening process was conducted using Rayyan [21], and any disagreements within a pair were resolved through discussion.

Studies were included if they applied a transformer-based language model to perform entity resolution, entity matching, record linkage, or data deduplication. Entity resolution is defined as the task of identifying and linking records, across one or more data sources,

that refer to the same real-world entity, a discrete, independently existing thing such as a person, organization, product, publication, or place [1,7].

Studies were excluded if they:

- Were not about entity resolution, entity matching, record linkage, or data deduplication as defined above;
- Addressed semantic textual similarity or duplicate-content detection (e.g., similar question detection, plagiarism detection) without matching records to a discrete, independently existing real-world entity;
- Utilized language models exclusively for blocking, without performing the actual matching or resolution;
- Did not apply a transformer-based language model (e.g., relied solely on rule-based methods, classical machine learning, or non-transformer deep learning architectures such as RNNs or CNNs);
- Only discussed the potential of language models for entity resolution without any systematic evaluation using standard accuracy metrics;
- Were published in any language other than English.

Studies that passed the title and abstract screening were retrieved for full-text review. The full-text screening was also conducted independently by the same reviewer pairs, with disagreements resolved through discussion.

2.3. Data Extraction and Synthesis

Data extraction was divided among four authors in a non-overlapping fashion, who extracted information from the assigned studies. Given the scale of the review, extraction was not independently duplicated; any uncertainties or ambiguous cases encountered during extraction were discussed among the team and resolved collectively. A structured data extraction template was developed to systematically collect relevant information from the included studies. The extracted data included the following key aspects:

- **Entity Focus:** The type of real-world entity the study aimed to link.
- **Blocking Approach:** Any pre-processing used to reduce candidate pairs before matching.
- **Matching Approach:** The methodology used for matching, including supervised and unsupervised methods.
- **Models Evaluated:** All transformer-based language models of study were extracted directly from their performance tables and classified by architecture type. Where a study reported results under a custom or author-assigned name, the model's original checkpoint name was recorded instead, to maintain consistency across studies and enable synthesis. Architecture type (e.g., encoder-only, decoder-only, encoder–decoder) was then derived from each model's checkpoint name based on publicly available information at the time of synthesis.
- **Learning Setting:** How each model was adapted for the task (e.g., zero-shot prompting, fine-tuning).
- **Benchmark Datasets:** All the dataset names that a study used for evaluating the language models.
- **Benchmark Entity Types:** The real-world entity type represented by each benchmark dataset that a study used.
- **Performance Metrics:** For each dataset a study evaluated on, only the metrics of the best-performing model–learning setting combination on that dataset were extracted. Because entity resolution is fundamentally a binary match/no-match classification task, precision, recall, and F1-score were extracted for the positive (match) class specifically, consistent with standard reporting conventions in the entity resolution literature, where the substantial class imbalance between matching and non-matching

record pairs makes match-class metrics more informative than aggregate (e.g., macro-averaged) alternatives; accuracy was also extracted where reported. F1-score, the harmonic mean of precision and recall [22], was the primary metric used for synthesis, as it was the most consistently reported metric across the included studies. In cases where only positive-class precision and recall were reported, positive-class F1-score was computed accordingly, as:

$$F_1 = \frac{2 \times (\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}}$$

Studies with neither a reported nor a calculable positive-class F1-score were excluded from the performance-related analyses; this included a small number of studies that reported only an aggregate F1-score (e.g., macro-, micro-, or weighted-averaged across classes) without a positive-class value, as such aggregate scores are not directly comparable to the positive-class F1-scores used throughout the rest of the corpus. For studies that evaluated both blocking and matching, only the performance metrics of the matching stage were extracted.

3. Results

This section reports the results of the review. It begins with the study selection process and the characteristics of the included studies, followed by publication trends and entity focus, blocking and matching strategies, the language models and architectures evaluated, benchmark dataset usage, and reported performance across the included studies.

3.1. Study Selection

A total of 1481 papers were retrieved from the database searches. After removing 640 duplicate records, 841 unique studies remained. These studies were screened based on their title and abstract, resulting in the exclusion of 646 articles. The remaining 195 studies proceeded to full-text screening, where an additional 40 studies were excluded. Ultimately, 155 studies were included for data extraction and synthesis. The PRISMA flow diagram in Figure 1 summarizes the entire process.

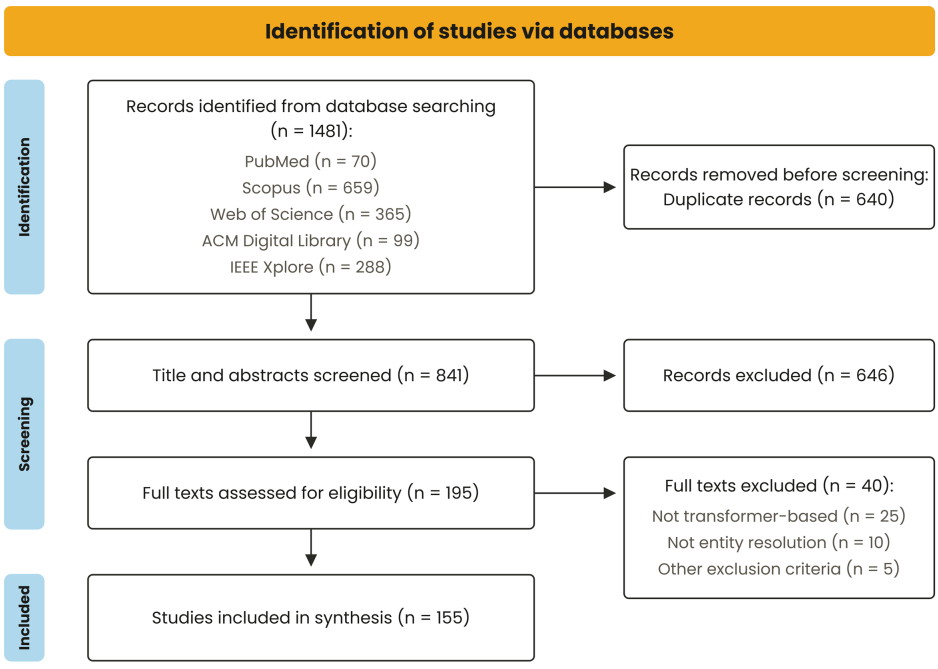

Figure 1. PRISMA flow diagram summarizing the screening process.

3.2. Characteristics of Included Studies

Table 1 summarizes the key characteristics of the 155 included studies, reporting the entity focus, blocking approach, and matching approach for each study. Blocking approaches are abbreviated as EBSS (Embedding-Based Semantic Similarity) and SBS (String-Based Similarity/Rule-Based); matching approaches are abbreviated as EB-CH (Embedding-Based with Classification Head), EB-ST (Embedding-Based with Similarity/Distance Threshold), EB-CG (Embedding-Based with Clustering or Graph-Based Grouping), and PB (Prompt-Based/Generative). For studies that employed more than one blocking or matching approach, all reported approaches are listed in the table, separated by a vertical bar (|).

Table 1. Key characteristics of the included studies.

Ref.	Year	Entity Focus	Blocking	Matching
[15]	2020	General purpose	Hybrid	EB-CH
[23]	2020	General purpose	-	EB-CH
[24]	2020	Knowledge-graph/Ontology	EBSS	EB-CG
[25]	2020	General purpose	-	EB-CH
[26]	2021	General purpose	-	EB-CH
[27]	2021	Product/E-commerce	-	EB-CH
[28]	2021	General purpose	-	EB-CH
[29]	2021	General purpose	EBSS	EB-CH
[30]	2021	General purpose	SBS	EB-CH
[31]	2021	General purpose	EBSS	EB-CH
[32]	2021	Music	SBS	EB-CH
[33]	2021	Bibliographic	-	EB-CH
[16]	2022	General purpose	-	EB-CH | EB-ST
[34]	2022	Geospatial/Address	Hybrid	EB-CH
[35]	2022	General purpose	-	EB-CH
[36]	2022	General purpose	SBS	EB-CH
[37]	2022	Product/E-commerce	-	EB-CH
[38]	2022	Product/E-commerce	-	EB-CH
[39]	2022	Product/E-commerce	-	EB-CH
[40]	2022	General purpose	-	EB-CH
[41]	2022	General purpose	SBS	EB-ST
[42]	2022	Product/E-commerce	-	EB-CH
[43]	2022	General purpose	-	EB-CH
[44]	2023	Product/E-commerce	-	EB-CH
[45]	2023	General purpose	-	EB-CH
[46]	2023	General purpose	-	EB-CH
[47]	2023	General purpose	EBSS	EB-CG | EB-CH
[48]	2023	General purpose	-	EB-CH
[49]	2023	General purpose	-	PB
[50]	2023	General purpose	-	EB-CH
[51]	2023	General purpose	-	EB-CH
[52]	2023	Person/Identity	SBS	EB-CH
[53]	2023	Bibliographic	SBS	EB-CH
[54]	2023	General purpose	Hybrid	EB-CH
[55]	2023	General purpose	-	EB-CH
[56]	2023	General purpose	-	PB | EB-CH
[57]	2023	General purpose	-	EB-CH
[58]	2023	Geospatial/Address	-	EB-CH
[59]	2023	General purpose	-	EB-CH
[60]	2024	Geospatial/Address	-	EB-CH | PB
[61]	2024	General purpose	EBSS | SBS	EB-CH | PB
[62]	2024	General purpose	-	PB
[63]	2024	General purpose	EBSS	EB-CH
[64]	2024	General purpose	-	EB-ST
[65]	2024	General purpose	-	EB-CH
[66]	2024	General purpose	-	EB-CH
[67]	2024	General purpose	-	Hybrid
[68]	2024	General purpose	EBSS	EB-ST | EB-CH | PB
[69]	2024	General purpose	-	EB-CH
[70]	2024	General purpose	-	EB-CH
[71]	2024	General purpose	-	EB-CH

Table 1. *Cont.*

Ref.	Year	Entity Focus	Blocking	Matching
[72]	2024	General purpose	-	EB-CH
[73]	2024	General purpose	-	EB-ST | PB
[74]	2024	General purpose	-	EB-CH
[75]	2024	Product/E-commerce	-	EB-CH
[76]	2024	General purpose	-	EB-CH
[77]	2024	Organization/Business	Hybrid	EB-CH
[78]	2024	Product/E-commerce	-	PB
[79]	2024	General purpose	-	EB-CH
[80]	2024	General purpose	-	EB-CH | PB
[81]	2024	Product/E-commerce	-	PB
[82]	2024	Geospatial/Address	EBSS	EB-CH | PB
[83]	2024	General purpose	SBS	EB-CH
[84]	2024	General purpose	EBSS	EB-ST
[85]	2024	General purpose	-	EB-CH | PB
[86]	2024	Geospatial/Address	Hybrid	EB-CH | PB
[87]	2024	General purpose	-	PB | EB-CH
[88]	2024	Geospatial/Address	-	EB-CH | PB
[89]	2024	Person/Identity	EBSS | SBS	EB-CH | PB
[90]	2024	Product/E-commerce	-	EB-CH
[91]	2025	General purpose	Hybrid	PB
[92]	2025	General purpose	-	PB
[93]	2025	General purpose	SBS	PB
[94]	2025	Industrial/Technical	-	PB
[95]	2025	Person/Identity	EBSS	EB-CH
[96]	2025	General purpose	EBSS	EB-ST
[97]	2025	General purpose	-	EB-CH
[98]	2025	General purpose	-	EB-CH | PB
[99]	2025	Person/Identity	EBSS	Hybrid
[100]	2025	Product/E-commerce	-	PB
[101]	2025	General purpose	-	EB-ST | PB
[102]	2025	General purpose	EBSS	EB-ST | EB-CH
[103]	2025	Product/E-commerce	EBSS	EB-CH
[104]	2025	Person/Identity	-	EB-ST
[105]	2025	General purpose	Hybrid	EB-CH
[106]	2025	General purpose	-	EB-CH
[107]	2025	Industrial/Technical	EBSS	PB | EB-CH
[108]	2025	General purpose	EBSS	EB-CH | PB
[109]	2025	General purpose	-	EB-CH | EB-ST
[110]	2025	General purpose	SBS	PB
[111]	2025	General purpose	-	EB-ST
[112]	2025	Product/E-commerce	SBS	EB-CH
[113]	2025	General purpose	EBSS	PB
[114]	2025	General purpose	SBS	EB-CH | PB
[115]	2025	Product/E-commerce	EBSS	EB-ST | EB-CH | PB
[116]	2025	General purpose	SBS	EB-CH
[117]	2025	General purpose	SBS	PB | EB-CH
[118]	2025	Knowledge-graph/Ontology	EBSS	Hybrid
[119]	2025	General purpose	SBS	PB | EB-CH
[120]	2025	Product/E-commerce	-	EB-CH
[121]	2025	General purpose	EBSS	PB | EB-CH
[122]	2025	Geospatial/Address	SBS	PB | EB-CH
[123]	2025	General purpose	-	PB | EB-CH
[124]	2025	Knowledge-graph/Ontology	-	PB
[125]	2025	Geospatial/Address	SBS	EB-CH
[126]	2025	Geospatial/Address	Hybrid	EB-CH
[127]	2025	General purpose	EBSS	EB-CG | EB-CH
[128]	2025	General purpose	-	PB
[129]	2025	General purpose	SBS	Hybrid
[130]	2025	General purpose	-	EB-CH
[131]	2025	General purpose	SBS	EB-CH
[132]	2025	General purpose	-	EB-CH | PB
[133]	2025	General purpose	-	EB-CH
[134]	2025	General purpose	-	PB
[135]	2025	General purpose	-	EB-CH
[136]	2025	Person/Identity	EBSS	Hybrid | EB-ST

Table 1. *Cont.*

Ref.	Year	Entity Focus	Blocking	Matching
[137]	2025	Person/Identity	EBSS	Hybrid | EB-CH
[138]	2025	Geospatial/Address	-	EB-CH
[139]	2025	Person/Identity	-	Hybrid
[140]	2025	General purpose	-	EB-CH
[141]	2026	General purpose	Hybrid	PB
[142]	2026	Product/E-commerce	-	EB-CH | PB
[143]	2026	General purpose	EBSS	Hybrid
[144]	2026	General purpose	-	PB | EB-CH
[145]	2026	General purpose	EBSS	Hybrid
[146]	2026	General purpose	-	PB
[147]	2026	Industrial/Technical	-	PB | Hybrid
[148]	2026	General purpose	-	EB-CH | EB-ST
[149]	2026	Product/E-commerce	-	EB-CH
[150]	2026	Product/E-commerce	-	Hybrid
[151]	2026	Industrial/Technical	-	EB-CH
[152]	2026	Person/Identity	SBS	Hybrid
[153]	2026	Person/Identity	-	PB
[154]	2026	General purpose	-	EB-CH
[155]	2026	Organization/Business	EBSS	PB
[156]	2026	General purpose	EBSS	PB
[157]	2026	Product/E-commerce	EBSS	Hybrid
[158]	2026	General purpose	-	EB-CH | PB
[159]	2026	General purpose	EBSS	EB-CH
[160]	2026	General purpose	-	EB-CH | PB
[161]	2026	General purpose	-	EB-CH
[162]	2026	Person/Identity	-	PB
[163]	2026	Geospatial/Address	-	EB-CH | PB
[164]	2026	General purpose	SBS	EB-ST
[165]	2026	General purpose	-	EB-CH | PB
[166]	2026	General purpose	SBS	EB-CH
[167]	2026	General purpose	EBSS	EB-CG | EB-CH | PB
[168]	2026	General purpose	-	EB-CH | PB
[169]	2026	Person/Identity	-	EB-CH
[170]	2026	Organization/Business	-	Hybrid
[171]	2026	Organization/Business	-	Hybrid
[172]	2026	General purpose	SBS	PB
[173]	2026	General purpose	SBS	PB
[174]	2026	General purpose	-	PB | EB-CH
[175]	2026	General purpose	-	EB-CH

3.3. Publication Trends and Entity Focus

The evolution of research on transformer-based language models for entity resolution, both over time and across different entity types, provides a broader view of the development of this field. The literature has grown steadily and substantially over the review period (Figure 2a). Included studies rose from four in 2020 (3%) to eight in 2021 (5%) and eleven in 2022 (7%), before accelerating sharply to sixteen in 2023 (10%), thirty-one in 2024 (20%), and fifty in 2025 (32%), the single largest annual share observed. Thirty-five studies (23%) were published in 2026 at the time of the search (17 August 2026); as this represents roughly seven and a half months of activity rather than a full year, the corresponding bar understates 2026 output.

Early in the review period (2020–2021), included studies were concentrated in general-purpose matching, with one study each addressing knowledge-graph/ontology alignment [24], product/e-commerce [27], bibliographic [33], and music matching [32]. From 2022 onward, the range of represented domains broadened to include geospatial/address [34], person/identity [52], organization/business [77], and industrial/technical matching [94], alongside continued growth in general-purpose studies. The entity focus of every included study, by publication year, is provided in Table 1.

Considered in aggregate across the full review period, general-purpose entity matching accounts for 64% of the 155 included studies (99/155), the single largest entity-focus category (Figure 2b). Product/e-commerce matching is the next most common at 12% (19/155), followed by person/identity resolution at 8% (12/155) and geospatial/address matching at 7% (11/155). The remaining five categories (organization/business matching, 3%, 4/155; industrial/technical matching, 3%, 4/155; knowledge-graph/ontology alignment, 2%, 3/155; bibliographic matching, 1%, 2/155; and music matching, 1%, 1/155) together account for the remaining 9% of included studies.

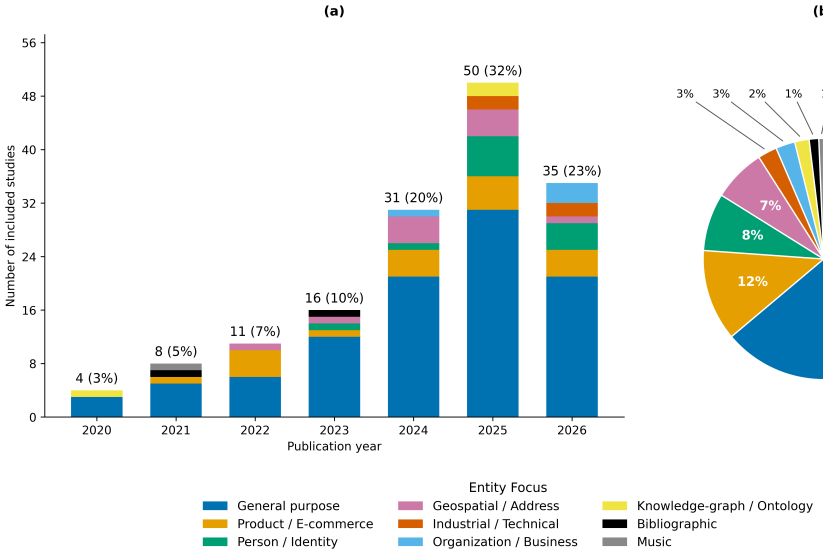

Figure 2. (**a**) Number of included studies by publication year, stratified by entity focus. The 2026 count reflects a partial year, through the search date of 17 August 2026. (**b**) Overall entity-focus distribution across all 155 included studies.

3.4. Blocking and Matching Strategies

Entity resolution pipelines in this literature are typically built from two sequential stages: a blocking step that narrows the candidate comparison space, and a matching step that renders the final decision. The review captured which approach each included study used at both stages.

The majority of included studies, 92 of 155 (59%), did not perform a blocking step of their own; instead, they evaluated directly on candidate pairs already provided by an established benchmark or external data source, or performed an exhaustive (unfiltered) comparison across the full dataset (Figure 3a). Among the remaining studies, 31 (20%) applied embedding-based semantic similarity for blocking, 25 (16%) applied string-based or rule-based blocking, and 9 (6%) applied a hybrid approach combining multiple blocking mechanisms; because a study may report more than one blocking approach, these counts sum to more than 155. The share of studies that did not perform a blocking step of their own ranged from 44% to 75% across individual years of the review period, without a consistent directional trend over time (Figure 3b).

For the final matching decision, 109 of 155 studies (70%) used an embedding-based approach with a classification head, the most common mechanism overall (Figure 3c). Prompt-based or generative matching, in which the language model directly outputs the match decision, was used by 56 studies (36%); embedding-based matching with a similarity or distance threshold was used by 16 studies (10%); hybrid matching mechanisms combining multiple approaches were used by 15 studies (10%); and embedding-based matching with clustering or graph-based grouping was used by 4 studies (3%); as with

blocking, a study may report more than one matching mechanism, so these counts also sum to more than 155. Classification-head usage declined steadily from 100% in 2021 to 51% in 2026, while prompt-based/generative matching, absent before 2023, rose over the same period to reach parity at 51% in 2026 (Figure 3d). The blocking and matching approach used by each individual study is provided in Table 1.

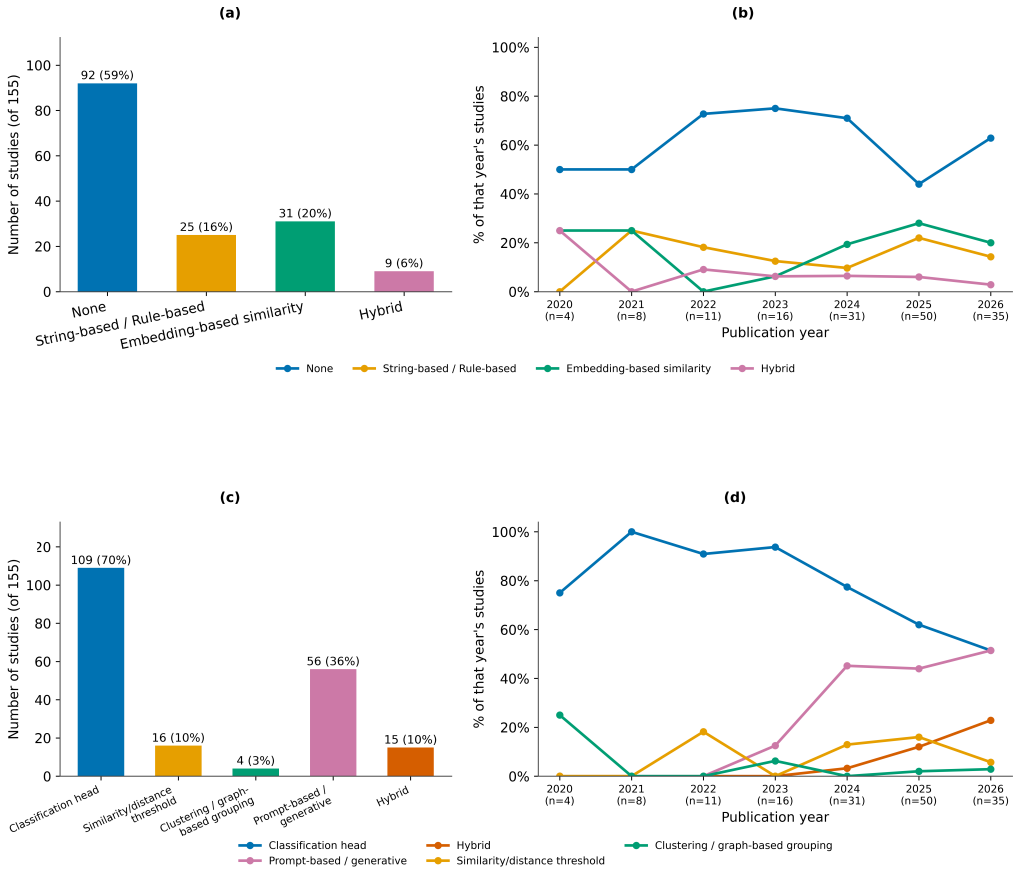

Figure 3. (**a**) Number of included studies by blocking approach used prior to matching. (**b**) Share of studies using each blocking approach, by publication year. (**c**) Number of included studies by matching approach used for the final decision. (**d**) Share of studies using each matching approach, by publication year.

3.5. Language Models, Architecture Evaluated, and Learning Setting

In this subsection, we synthesize how the specific language models used to evaluate entity resolution pipelines varied substantially across studies. A total of 205 distinct language models were evaluated across the 155 included studies. RoBERTa was the most frequently evaluated model, used in 71 studies (46%), followed by BERT in 55 studies (35%) and DistilBERT in 21 studies (14%) (Figure 4). Among generative models, GPT-4 was the most commonly evaluated, appearing in 13 studies (8%), followed by GPT-3.5-turbo (12 studies, 8%) and GPT-4o-mini (11 studies, 7%), and Mistral-7B (9 studies, 6%). The remaining 190 distinct models were each evaluated in a smaller number of studies, together accounting for 262 study-model pairs.

In terms of model architecture, encoder-only models were evaluated in 128 studies (83%), the most common architecture type overall, followed by decoder-only models in 65 studies (42%). Encoder–decoder architectures were evaluated in 7 studies (5%), and ensemble combinations of multiple architecture types in 5 studies (3%) (Figure 5a); because a

study may evaluate models of more than one architecture type, these categories are not mutually exclusive and do not sum to 155. The composition of evaluated architectures shifted over the review period: encoder-only models were evaluated in 100% of studies from 2020 through 2022, while decoder-only models were entirely absent from included studies over the same years. The earliest evaluations of decoder-only architectures among the included studies appeared in 2023 [44,49,56], preceding the broader shift visible in Figure 5b: from 2023 onward, decoder-only models were evaluated with increasing frequency, becoming the second most commonly evaluated architecture type, while encoder-only models retained a reduced but continued majority share in 2025 and 2026. The complete list of models and architecture types evaluated by each individual study is provided in Supplementary Table S1.

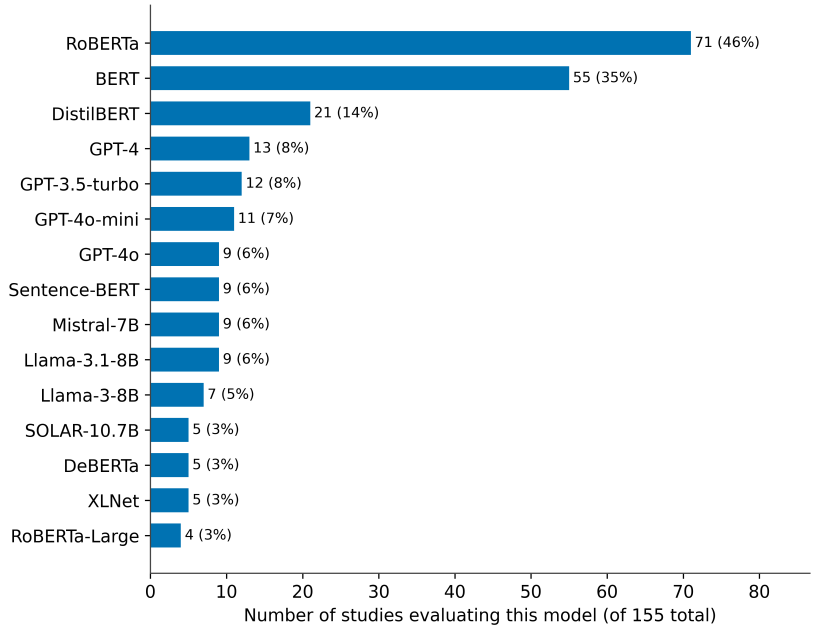

Figure 4. The 15 most frequently evaluated language models across the 155 included studies, out of 205 distinct models in total.

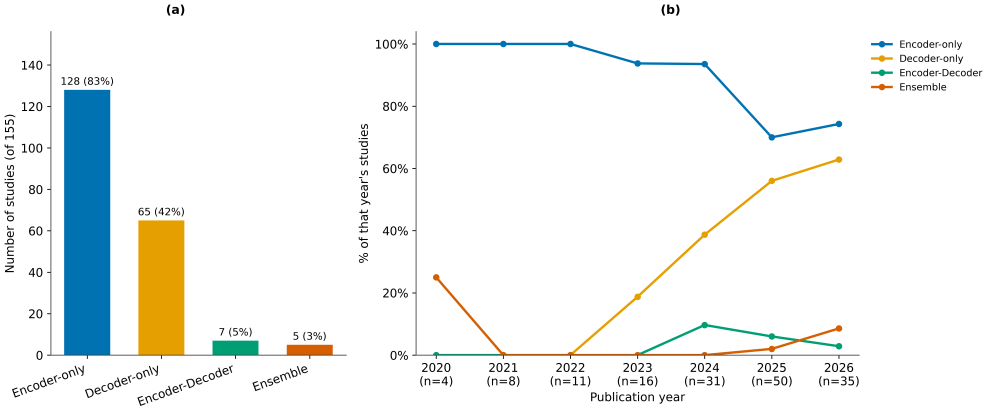

Figure 5. (**a**) Number of included studies evaluating each model architecture type. (**b**) Share of studies evaluating each architecture type, by publication year. A study may evaluate models of more than one architecture type.

In addition to architecture type, the review captured the learning setting used to adapt each model for entity resolution: whether it was fine-tuned, used with frozen embeddings, or prompted directly. Full parameter fine-tuning remained the dominant approach, used in 120 studies (77%), though zero-shot prompting (54 studies, 35%) and few-shot prompting

(30 studies, 19%) were also common. Feature-based evaluation using frozen embeddings without fine-tuning was used in 29 studies (19%), and parameter-efficient fine-tuning (PEFT) in 22 studies (14%) (Figure 6); because a study may evaluate more than one learning setting, these counts do not sum to 155.

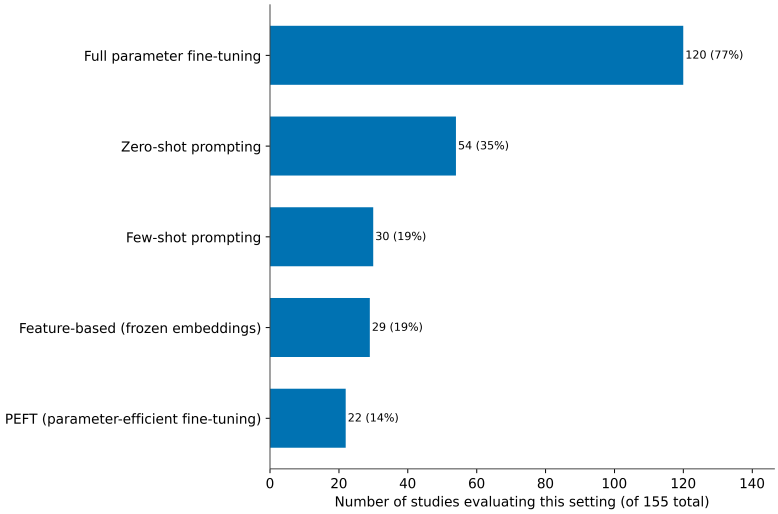

Figure 6. Learning settings evaluated across the 155 included studies. A study may evaluate more than one learning setting.

3.6. Benchmark Dataset Usage and Concentration

Across the 155 included studies, evaluation datasets were concentrated in a small number of entity types. Product/e-commerce datasets were used by 111 studies (72%), the most commonly represented entity type in evaluation data, followed by bibliographic/publication datasets in 77 studies (50%) and music datasets in 45 studies (29%) (Figure 7). Restaurant datasets were used by 33 studies (21%), person/identity datasets by 24 studies (15%), geospatial/place datasets by 21 studies (14%), organization/business datasets by 20 studies (13%), and movie/TV media datasets by 19 studies (12%). Knowledge-graph/ontology datasets and industrial/technical datasets were each used by 5 studies (3%). The complete list of benchmark datasets evaluated by each individual study is provided in Supplementary Table S1.

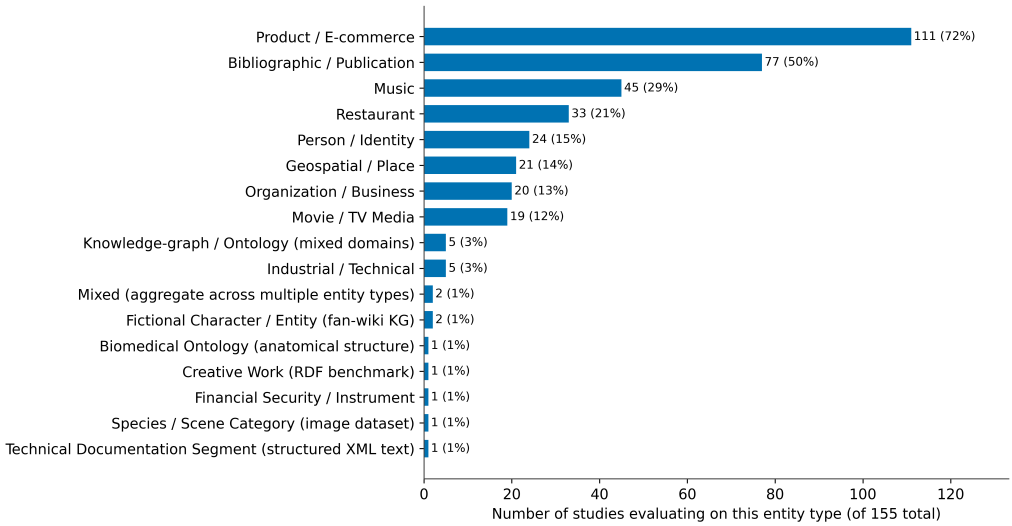

Figure 7. Real-world entity types represented in the evaluation datasets used across the 155 included studies. A study may evaluate on datasets spanning more than one entity type.

For each general-purpose study, the number of distinct real-world entity types represented among its evaluation datasets was counted (Figure 8): 34 studies (34%) evaluated on three distinct entity types, the most common pattern, followed by 21 studies (21%) on four and 17 studies (17%) on two. Fourteen studies (14%) evaluated on only a single distinct entity type [28,56,128], and 12 studies (12%) and 1 study (1%) evaluated on five and six distinct entity types, respectively. In aggregate, 31 of the 99 general-purpose studies (31%) evaluated on two or fewer distinct entity types.

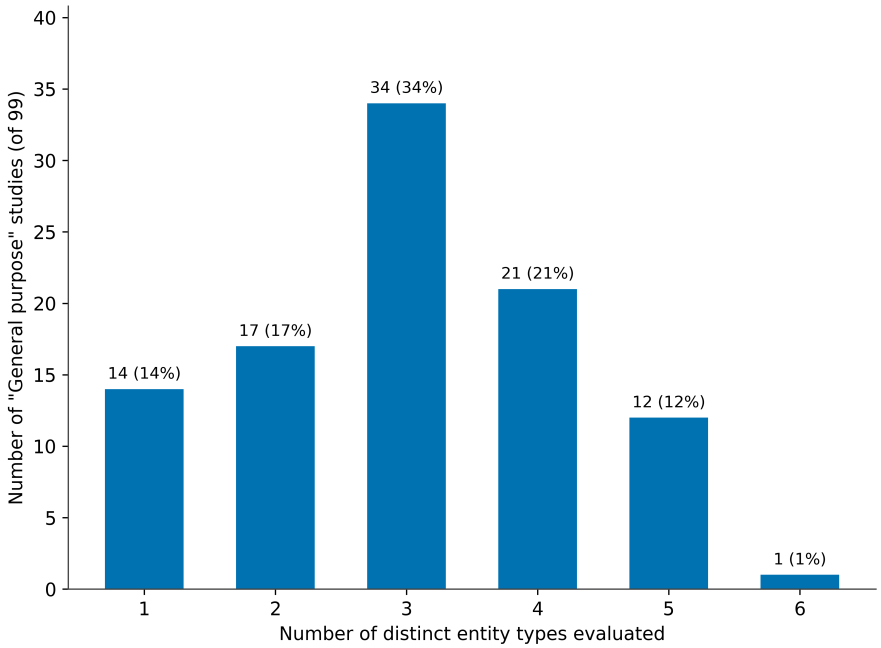

Figure 8. Distribution of the 99 general-purpose studies by the number of distinct real-world entity types represented among each study's evaluation datasets.

3.7. Reported Performance

After capturing the models, architectures, and benchmark datasets in each study, we compared their reported performance across datasets and architectures.

Best-per-study mean F1 ranged from 77.2 to 97.6 across the 15 most frequently used benchmark datasets (Figure 9). Reported performance was highest on DBLP–ACM Dirty (n = 21, mean F1 = 97.6), Fodors–Zagats (n = 24, mean F1 = 97.2), and DBLP–ACM Structured (n = 47, mean F1 = 96.8). Performance on Amazon–Google (n = 63), the most frequently used dataset among the included studies, was comparatively low and highly variable, with a mean F1 of 77.2 and a wide spread that included a number of low-performing outliers. The Walmart–Amazon Dirty (n = 24, mean F1 = 84.9) and Structured (n = 56, mean F1 = 85.0) variants showed similarly moderate performance. Abt–Buy (n = 52, mean F1 = 88.1), Beer (n = 30, mean F1 = 92.3), DBLP–Scholar Dirty (n = 22, mean F1 = 93.3) and Structured (n = 45, mean F1 = 93.8), and iTunes–Amazon Dirty (n = 20, mean F1 = 91.4) and Structured (n = 31, mean F1 = 93.9) fell in an intermediate-to-high range. The remaining 2 of the 15 most frequently used datasets, Company (n = 12, mean F1 = 90.8) and the WDC computers benchmark, evaluated as three separate size-based subsets in the underlying data (n = 11 each, mean F1 ranging from 88.5 to 95.8), also showed consistently high performance despite their smaller sample sizes. Best-per-study F1, aggregated by dataset, for the full set of benchmark datasets evaluated across the included studies is provided in Table A1.

Figure 9. Best-per-study F1 score for the 15 most frequently used benchmark datasets. Each point represents one study's single best-performing model/setting on that dataset.

Comparing model architectures on the same dataset shows no single architecture type that consistently outperformed the other across datasets (Figure 10). On Abt–Buy, the mean best-per-study F1 for decoder-only models (89.8, n = 13) exceeded that of encoder-only models (87.4, n = 38); on DBLP–ACM Structured, the reverse held, with encoder-only models achieving a higher mean F1 (97.4, n = 35) than decoder-only models (95.0, n = 11). On Amazon–Google and Walmart–Amazon Structured, encoder-only and decoder-only mean F1 scores were within roughly one point of each other. Across the datasets shown, decoder-only models were evaluated in substantially fewer studies than encoder-only models on every dataset; this smaller sample size, reflecting the more recent adoption of decoder-only architectures in this literature, means architecture comparisons should be interpreted as suggestive rather than conclusive, particularly for the lowest-n cells. Best-per-study F1 by architecture type, reported individually for each study and dataset, is provided in Supplementary Table S2.

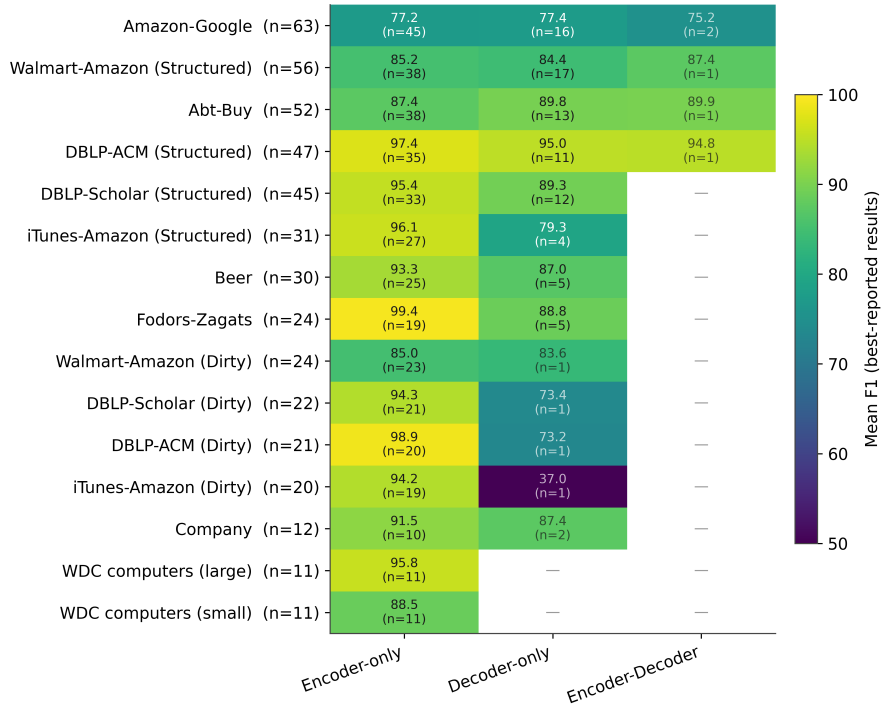

Figure 10. Mean best-per-study F1 score by model architecture type and benchmark dataset. Cells report the number of contributing studies (n); a dash indicates no study's best result on that dataset used that architecture.

4. Discussion

This systematic scoping review provides an updated overview of how transformer-based language models have been used for entity resolution. By examining the models, learning approaches, matching strategies, application areas, benchmark datasets, and reported performance across the included studies, this review brings together a large and rapidly growing body of research that has developed alongside recent advances in transformer-based methods. The findings provide a broad view of how the field has evolved and where the current evidence is strongest or still limited. In the following discussion, we interpret the main patterns identified in the review, consider their implications for entity resolution research and practice, and highlight areas where further work is needed.

4.1. Literature Growth and Domain Concentration

One of the clearest patterns identified in this review is the rapid growth of research on transformer-based approaches to entity resolution. More included studies were published in 2025 and the partial 2026 period combined than in all of the preceding years covered by the review, showing how quickly interest in this area has increased. This pattern is especially notable because the 2026 count reflects only studies identified up to the search cutoff, meaning that the total for the year is likely to increase further if the recent pace of research continues. This increase is consistent with the broader growth of research on generative AI and LLMs, where recent studies have documented a sharp rise in research activity across a wide range of fields, alongside rapid advances in the capabilities and applications of LLMs [14,176]. These developments have created new opportunities to apply language models to tasks that require understanding and comparing complex information, which may partly explain the growing interest in their use for entity resolution. As research in this area continues to grow, regularly reassessing the literature will be important for understanding how the field is changing and where important gaps remain.

Another important pattern is the strong concentration of studies on general-purpose entity matching, even as the range of application areas has broadened over time. Most of the included studies were designed as general-purpose approaches rather than for a specific type of real-world entity, while product and e-commerce matching represented the largest domain-specific area. Research on person and identity resolution, geospatial and address matching, organizations, and industrial applications was much less common. The appearance of these areas mainly in the later years of the review suggests that transformer-based entity resolution is gradually being applied to a wider range of practical problems. However, the evidence for broad generalizability remains limited. Among studies classified as general-purpose, nearly one-third were evaluated on only one or two distinct real-world entity types, while only a small proportion were tested across five or more. This suggests that presenting a framework or pipeline as broadly generalizable does not necessarily mean that its performance has been demonstrated across a wide range of entity types. The continued concentration of evaluation in a limited set of domains therefore raises questions about how well findings from commonly used benchmarks transfer to more specialized real-world settings, where data characteristics, error patterns, and matching requirements may differ considerably. This concern is also reflected in the benchmark datasets used across the literature, which were heavily concentrated in product/e-commerce, bibliographic, and music data. The repeated use of established benchmarks is valuable because it provides a common basis for comparing methods across studies. However, relying too heavily on the same datasets can limit what can be learned about performance under different data conditions and may encourage methods to be optimized around a relatively narrow set of well-studied benchmarks. As a result, high performance on commonly used datasets should not by itself be taken as evidence that a method will perform equally well in less represented domains. Broader evaluation across more diverse datasets, particularly those drawn from real-world and domain-specific applications, would provide stronger evidence of robustness and generalizability.

4.2. Model Architecture and Learning Strategies

Following the growth and widening application of transformer-based entity resolution, the review also shows a clear change in the types of model architectures being evaluated. Encoder-only models, particularly BERT- and RoBERTa-based approaches, dominated the earlier literature and remain widely used, reflecting their strong performance in representation learning and supervised classification tasks. Their continued use also shows that they remain important reference points for the field, with BERT- and RoBERTa-based models frequently serving as baselines against which newer models and approaches are evaluated. However, decoder-only models began to appear in the entity resolution literature from 2023 onward and have since been evaluated with increasing frequency. This change closely follows the broader development of generative LLMs, whose improved instruction-following, in-context learning, and reasoning capabilities have made it possible to approach entity matching through prompting rather than relying only on task-specific classification models [177,178]. The two architecture types offer different advantages and limitations. Encoder-only models are generally more computationally efficient for representation learning and classification and can be well suited to settings where large numbers of candidate record pairs must be processed. However, when used as supervised entity matching models, they typically require task-specific training and labeled examples to learn the matching decision. Creating high-quality labeled record pairs can be costly and time-consuming, particularly in domains where annotation requires substantial manual review or domain expertise. Their ability to transfer directly to a new task or domain without additional adaptation is therefore more limited than that of instruction-tuned decoder-only

models. Decoder-only models, in contrast, can use task instructions and examples provided at inference time, allowing them to perform entity matching in zero-shot or few-shot settings without training a separate classifier for each new task [87]. This flexibility can be particularly valuable when labeled data are scarce or when a model needs to be applied across different domains, although it often comes with higher computational and inference costs. These tradeoffs suggest that the choice between encoder-only and decoder-only models should consider not only predictive performance, but also the availability and cost of labeled data, the need for adaptation across tasks or domains, and the computational resources required for deployment.

These differences in model architecture are closely tied to how the models are adapted for entity resolution. Full-parameter fine-tuning remains the most commonly used learning setting, suggesting that task-specific training is still an important approach for achieving strong matching performance. At the same time, the growing use of zero-shot and few-shot prompting and parameter-efficient fine-tuning reflects increasing interest in reducing the amount of labeled data and computational effort required to adapt language models to new entity resolution tasks. This is particularly relevant given the cost of creating labeled training data, as discussed above. Approaches such as prompting can avoid task-specific training altogether, while parameter-efficient fine-tuning offers a middle ground by adapting only a small portion of the model rather than updating all parameters [179]. These alternatives can make it easier to apply language models across new datasets and domains, but they may also involve tradeoffs in consistency and performance compared with fully fine-tuned models. Overall, the range of learning settings observed in this review suggests that recent work is increasingly focused not only on improving matching accuracy, but also on finding more practical ways to adapt language models when labeled data or computational resources are limited.

4.3. Blocking and Matching Approaches

These changes in model architecture and learning strategies are also reflected in how the final matching decision is made. Earlier studies largely treated entity matching as a supervised classification problem, using transformer representations together with a classification head to determine whether two records refer to the same entity. Although this remains the most common matching approach overall, its dominance has declined as prompt-based matching has become increasingly common. Prompt-based approaches offer a different way of framing the task by allowing the language model to produce the match decision directly from instructions and record information, without requiring a separately trained classification head. The fact that prompt-based matching was absent from the earlier literature but reached the same share of studies as classification-head matching in 2026 shows how quickly this approach has gained attention. Rather than indicating that classification-based methods are being replaced, this pattern suggests that researchers now have a broader set of options for designing the matching stage. Classification-based approaches remain attractive when labeled data are available and efficient large-scale inference is important, while prompt-based approaches may be more useful when flexibility across tasks or limited supervision is a priority. The growing use of both approaches also reinforces the need for direct comparisons that consider not only matching accuracy, but also training requirements, inference cost, and performance across different domains.

A separate consideration is the extent to which matching approaches were evaluated within complete entity resolution pipelines that also included blocking. Most studies in this review did not implement their own blocking procedure and instead evaluated matching on candidate pairs provided by established benchmarks or external sources,

or used exhaustive comparisons. Evaluating the matching stage in this way is useful for isolating matcher performance and supporting controlled comparisons across methods, but it provides less evidence about how the same approaches would behave when candidate generation and matching are combined in an end-to-end system. The two stages can also influence one another in practice, since blocking determines which candidate pairs reach the matcher, while the computational demands of the matching model can affect how much the candidate space needs to be reduced. This interaction may be particularly important for computationally intensive language models, where applying the matcher to a very large number of candidate pairs can substantially increase inference cost.

4.4. Benchmark Comparisons

Reported performance across the included studies was generally strong on many of the commonly used benchmark datasets, but the results also revealed an important difference in how informative these benchmarks may be for comparing newer methods. Several datasets showed consistently high F1 scores across studies, with some results approaching the upper limit of the metric. When performance becomes concentrated near this level, further improvements may be difficult to distinguish, and the benchmark may provide less separation between competing approaches. In contrast, datasets such as Amazon–Google and Walmart–Amazon showed lower average performance and greater variation across studies, suggesting that they continue to provide more room for distinguishing differences among methods. The architecture-level comparison also showed no consistent advantage for either encoder-only or decoder-only models across datasets. Decoder-only models performed better on some benchmarks, encoder-only models performed better on others, and several differences were relatively small. These comparisons should be interpreted cautiously because decoder-only models were represented by fewer studies on every benchmark included in the architecture analysis. Taken together, the findings do not support a simple conclusion that newer decoder-only models provide a general performance advantage over established encoder-based approaches. They also highlight the importance of evaluating new methods on benchmarks that remain sufficiently challenging to meaningfully distinguish improvements in entity resolution performance.

4.5. Future Research Directions

Several areas identified in this review warrant greater attention in future research. A particularly notable gap in the literature is the limited application of transformer-based entity resolution to clinical and healthcare data. Although person and identity resolution were represented among the included studies, only one of these studies [152] applied a language model-based entity resolution approach to real-world patient data. Despite targeted efforts in this review to identify clinical and healthcare applications, evidence in this area remains extremely limited. This scarcity also extends beyond the peer-reviewed literature, with only limited preprint work examining the use of language models for patient record linkage, particularly using real-world healthcare data [180]. This gap is important because patient record linkage is widely used across healthcare and public health to connect information about the same individual across hospitals, laboratories, registries, and other data sources. Accurate linkage is therefore important for integrating fragmented health information and supporting longitudinal research, disease surveillance, and other population-level applications [3]. Clinical applications also present challenges that are not well represented by commonly used product or bibliographic benchmarks, including restricted access to identifiable data, privacy requirements, incomplete or inconsistent patient information, and the potentially greater consequences of incorrect matches [2]. These factors may partly explain the limited use of language model-based entity resolution

in healthcare, while also making it difficult to determine whether the strong performance reported on commonly used benchmarks will translate to real-world patient matching. Overall, the current evidence remains insufficient to draw broad conclusions about the effectiveness of transformer-based entity resolution in healthcare and public health settings.

Moreover, future research should place greater emphasis on evaluation across a wider range of entity types and datasets. Although many studies presented general-purpose entity resolution approaches, a substantial proportion were evaluated on only a small number of real-world entity types, while the overall literature remained heavily concentrated on product, bibliographic, and music datasets. Evaluating the same methods across more diverse domains would provide stronger evidence about their ability to generalize beyond the datasets on which they were developed or tuned. Greater use of real-world and domain-specific datasets would also help determine how language model-based approaches respond to different forms of missingness, noise, formatting variation, class imbalance, and record structure. In addition, several widely used benchmarks now show consistently high performance across studies, which may reduce their ability to distinguish meaningful improvements among newer approaches. Developing and adopting more challenging evaluation datasets could therefore provide a more informative basis for comparing future methods.

5. Limitations

This scoping review has several limitations that should be acknowledged.

Consistent with PRISMA-ScR guidance, under which formal critical appraisal of individual sources is optional for scoping reviews, this review did not conduct a methodological-quality assessment of the included studies [20]. Given the scale of the literature captured here, a systematic appraisal of reproducibility, dataset leakage, and evaluation-protocol validity across so many studies was not feasible, and reported F1-scores should not be read as an indicator of study quality.

Methodological heterogeneity also posed challenges during data synthesis. Variations in how studies defined and implemented key components such as blocking, matching strategies, and model configurations made it difficult to consistently categorize studies into clear thematic groups. As a result, some degree of abstraction and simplification was necessary, which may have led to the loss of finer-grained methodological details in favor of broader synthesis.

6. Conclusions

Transformer-based language models have expanded the range of approaches used for entity resolution, with growing use of prompting and few-shot prompting alongside established encoder-based methods. However, current evidence remains concentrated in a limited set of commonly used benchmarks. Studies designed as general-purpose approaches are also frequently evaluated on only a narrow range of entity types, raising questions about whether the available evidence supports broad generalizability. No model architecture showed a consistent performance advantage across the datasets examined, though these comparisons are limited by the smaller number of studies evaluating decoder-only models. Most studies also did not implement their own blocking step, instead evaluating matching on candidate pairs already provided by established benchmarks or external sources, which leaves limited evidence on how these methods perform within a complete, end-to-end pipeline. Applications to patient record linkage remain particularly scarce, despite their practical importance for healthcare and public health.

Future research should focus on generalizability, robustness, computational efficiency, and evaluation on diverse and challenging real-world datasets, particularly in under-

represented domains such as healthcare. By synthesizing reported performance across benchmarks, capturing the literature's continued growth following the emergence of generative large language models, and identifying the scarcity of research on clinical and healthcare applications, this review extends earlier surveys of the field. Future progress in transformer-based entity resolution will depend on demonstrating reliable performance beyond established benchmarks and under realistic deployment conditions.

Supplementary Materials: The following supporting information can be downloaded at https://www.mdpi.com/article/10.3390/info17090917/s1, Table S1: Complete list of models, architecture types, and benchmark datasets evaluated by each included study; Table S2: Best-per-study performance results by benchmark dataset.

Author Contributions: Conceptualization, M.B. and I.Z.; methodology, M.B. and M.S.; data curation, M.B., M.S., K.K. and T.M.R.K.; formal analysis, M.B., A.E.Z.S.A. and M.S.; resources, I.Z.; writing—original draft preparation, M.B. and M.S.; writing—review and editing, M.B., K.K., T.M.R.K., A.E.Z.S.A., M.S., S.A.B. and I.Z.; visualization, M.B. and A.E.Z.S.A.; supervision, I.Z.; funding acquisition, I.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Missouri Cancer Registry and Research Center, funded in part through a cooperative agreement between the Centers for Disease Control and Prevention and the Missouri Department of Health and Senior Services (MODHSS) (NU58DP007130-05), and through a surveillance contract between MODHSS and the University of Missouri. The ideas and opinions expressed are those of the author(s) and do not necessarily reflect the opinions of the MODHSS, the University of Missouri, CDC, or other funding agencies.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data extracted from the included studies, comprising the complete list of models, architecture types, and benchmark datasets evaluated by each study (Supplementary Table S1) and the corresponding best-per-study performance results by benchmark dataset (Supplementary Table S2), are provided as Supplementary Materials.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Full Search Strategies by Database

Appendix A.1. PubMed

("large language model*"[tiab] OR LLM[tiab] OR "language model*"[tiab] OR "transformer model*"[tiab] OR transformer[tiab] OR "generative AI"[tiab] OR "foundation model*"[tiab] OR "small language model*"[tiab] OR BERT[tiab] OR RoBERTa[tiab]) AND ("entity resolution"[tiab] OR "entity matching"[tiab] OR "identity resolution"[tiab] OR "record linkage"[tiab] OR "data linkage"[tiab] OR "probabilistic linkage"[tiab] OR "data deduplication"[tiab] OR "record deduplication"[tiab] OR deduplication[tiab] OR "duplicate detection"[tiab] OR "patient matching"[tiab] OR "person matching"[tiab] OR "master patient index"[tiab] OR "patient deduplication"[tiab]) AND ("2017"[pdat] : "2027"[pdat]) AND english[la]

The "Exclude preprints" filter (Additional Filters > Other) was applied to remove NIH Preprint Pilot content (medRxiv, bioRxiv, Research Square).

Appendix A.2. Scopus

TITLE-ABS-KEY("large language model*" OR LLM OR "language model*" OR "transformer model*" OR transformer OR "generative AI" OR "foundation model*" OR "small language model*" OR BERT OR RoBERTa) AND TITLE-ABS-KEY("entity resolution" OR "entity matching" OR "identity resolution" OR

"record linkage" OR "data linkage" OR "probabilistic linkage" OR "data deduplication" OR "record deduplication" OR deduplication OR "duplicate detection" OR "patient matching" OR "person matching" OR "master patient index" OR "patient deduplication") AND PUBYEAR > 2016 AND LANGUAGE(english)

The Document Type filter was restricted to Article, Conference Paper, and Review.

Appendix A.3. Web of Science

TS=("large language model*" OR LLM OR "language model*" OR "transformer model*" OR transformer OR "generative AI" OR "foundation model*" OR "small language model*" OR BERT OR RoBERTa) AND TS=("entity resolution" OR "entity matching" OR "identity resolution" OR "record linkage" OR "data linkage" OR "probabilistic linkage" OR "data deduplication" OR "record deduplication" OR deduplication OR "duplicate detection" OR "patient matching" OR "person matching" OR "master patient index" OR "patient deduplication") AND PY=2017-2027 AND LA=English

The Preprint Citation Index was excluded from the set of searched Web of Science databases.

Appendix A.4. ACM Digital Library

((Title:("large language model" OR "large language models" OR LLM OR "language model" OR "language models" OR "transformer model" OR "transformer models" OR transformer OR "generative AI" OR "foundation model" OR "foundation models" OR "small language model" OR "small language models" OR BERT OR RoBERTa)) OR (Abstract:("large language model" OR "large language models" OR LLM OR "language model" OR "language models" OR "transformer model" OR "transformer models" OR transformer OR "generative AI" OR "foundation model" OR "foundation models" OR "small language model" OR "small language models" OR BERT OR RoBERTa))) AND ((Title:("entity resolution" OR "entity matching" OR "identity resolution" OR "record linkage" OR "data linkage" OR "probabilistic linkage" OR "data deduplication" OR "record deduplication" OR deduplication OR "duplicate detection" OR "patient matching" OR "person matching" OR "master patient index" OR "patient deduplication")) OR (Abstract:("entity resolution" OR "entity matching" OR "identity resolution" OR "record linkage" OR "data linkage" OR "probabilistic linkage" OR "data deduplication" OR "record deduplication" OR deduplication OR "duplicate detection" OR "patient matching" OR "person matching" OR "master patient index" OR "patient deduplication")))

The Publication Date facet was restricted to 1 January 2017 through the search execution date; Content Type was restricted to Research Article.

Appendix A.5. IEEE Xplore

("large language model*" OR LLM OR "language model*" OR "transformer model*" OR transformer OR "generative AI" OR "foundation model*" OR "small language model*" OR BERT OR RoBERTa) AND ("entity resolution" OR "entity matching" OR "identity resolution" OR "record linkage" OR "data linkage" OR "probabilistic linkage" OR "data deduplication" OR "record deduplication" OR deduplication OR "duplicate detection" OR "patient matching" OR "person matching" OR "master patient index" OR "patient deduplication")

The Year filter was applied in the results sidebar (2017–present).

Appendix B. Summary Statistics of F1-Score by Benchmark Dataset

Table A1. Summary statistics for best-per-study F1-score (%), for all benchmark datasets used by more than one of the 155 included studies (datasets used by only a single study are omitted). N: number of contributing studies; Q1/Q3: first/third quartile; SD: standard deviation.

Dataset	N	Min	Q1	Median	Mean	Q3	Max	Range	SD
Amazon–Google	63	59.7	74.1	76.2	77.17	80.48	99.79	40.09	8.07
Walmart–Amazon (Structured)	56	52.2	84.7	86.76	84.99	89.4	100	47.8	9.8
Abt–Buy	52	44.93	89.29	90.9	88.06	93	98.94	54.01	10.29
DBLP–ACM (Structured)	47	56.47	97.9	99	96.79	99.15	99.89	43.42	7.58
DBLP–Scholar (Structured)	45	70	94.3	95.6	93.76	96	99.81	29.81	5.42
iTunes–Amazon (Structured)	31	37.63	94.6	97.06	93.91	98.18	100	62.37	11.5
Beer	30	55.43	92.82	94.37	92.27	96.55	100	44.57	8.54
Fodors–Zagats	24	44.25	100	100	97.15	100	100	55.75	11.43
Walmart–Amazon (Dirty)	24	65.74	84.35	85.69	84.92	87.16	95.89	30.15	5.75
DBLP–Scholar (Dirty)	22	73.44	92.65	95.6	93.31	95.75	98.65	25.21	5.56
DBLP–ACM (Dirty)	21	73.22	98.7	99.03	97.64	99.1	99.78	26.56	5.64
iTunes–Amazon (Dirty)	20	37	94.42	94.74	91.37	95.84	99.4	62.4	13.96
Company	12	80.73	87.96	92.82	90.79	93.85	93.85	13.12	4.08
WDC computers (small)	11	80.76	86.18	86.95	88.52	91.1	95.21	14.45	4.22
WDC computers (medium)	11	88.62	89.9	93.03	92.64	94.34	98.5	9.88	3
WDC computers (large)	11	91.7	94.78	95.95	95.83	97.23	98.5	6.8	2.08
WDC cameras (small)	10	80.89	84.64	86.65	86.84	89.94	92.25	11.36	3.74
WDC cameras (medium)	10	88.09	88.99	91.05	91.18	93.04	94.81	6.72	2.53
WDC cameras (large)	10	91.23	92.85	95.15	94.71	96.51	97.84	6.61	2.36
WDC computers (xlarge)	10	95.45	96.54	97.4	97.1	97.6	98.44	2.99	1.05
WDC shoes (small)	10	75.89	79.47	80.66	81.18	82.74	88.59	12.7	3.6
WDC shoes (medium)	10	82.66	83.51	86.22	86.3	87.75	93.73	11.07	3.33
WDC shoes (large)	10	87.37	91.29	93.32	92.82	95.33	97.16	9.79	3.4

Table A1. *Cont.*

Dataset	N	Min	Q1	Median	Mean	Q3	Max	Range	SD
WDC watches (small)	10	85.12	87.58	90.42	90.25	91.94	96.71	11.59	3.41
WDC watches (medium)	10	91.12	92.44	93.15	93.4	94.52	96.19	5.07	1.77
WDC watches (large)	10	93.62	96.48	97.12	97.11	98.46	98.97	5.35	1.63
WDC cameras (xlarge)	9	93.78	94.82	95.59	96.3	98.02	99.16	5.38	2.03
WDC shoes (xlarge)	9	90.11	90.79	96.63	94.71	97.88	98.47	8.36	3.67
WDC watches (xlarge)	9	96.53	96.66	97.09	97.26	97.61	99.11	2.58	0.83
Semi-REL	7	91.1	94.38	96.2	95.78	97.45	99.5	8.4	2.8
WDC Products	7	57.71	73.62	86.8	80.48	89.4	92.8	35.09	13.8
Rel-TEXT	6	60.7	62.39	65.6	68.11	69.3	84.9	24.2	8.94
Semi-HETER	6	69.7	78.5	92.8	87.45	96.6	97.8	28.1	12.43
Semi-TEXT-W	6	42.93	66.58	70.44	67.56	75.89	78.59	35.66	13.06
Cora	5	71	82.06	86.34	85.4	90	97.6	26.6	9.87
Semi-HOMO	5	93.8	94.1	94.3	94.74	95.2	96.3	2.5	1.02
Semi-TEXT-C	5	70.4	76.42	78.01	81.19	88.6	92.5	22.1	9.11
IMDB-TMDB	4	93.29	95.88	96.87	96.34	97.34	98.35	5.06	2.16
IMDB-TVDB	4	84.15	84.16	86.08	86.78	88.7	90.82	6.67	3.25
TMDB-TVDB	4	86.83	89.94	90.97	90.19	91.23	92	5.17	2.29
WDC Shoes	4	73.48	84.54	88.26	85.05	88.77	90.2	16.72	7.77
GEO-HETER	3	92.6	93	93.4	93.53	94	94.6	2	1.01
OSM-FSQ—Edinburgh	3	83.6	89.53	95.46	91.59	95.58	95.7	12.1	6.92
OSM-FSQ—Pittsburgh	3	83.6	88.15	92.7	90.07	93.3	93.9	10.3	5.63
OSM-FSQ—Singapore	3	80.4	84.9	89.4	86.53	89.6	89.8	9.4	5.32
OSM-FSQ—Toronto	3	82.6	88.6	94.6	90.71	94.76	94.92	12.32	7.02
OSM-Yelp—Edinburgh	3	90.8	93.95	97.1	95.16	97.34	97.58	6.78	3.78
OSM-Yelp—Pittsburgh	3	89	93.06	97.11	94.57	97.35	97.6	8.6	4.83
OSM-Yelp—Singapore	3	83	87.72	92.45	89.45	92.68	92.9	9.9	5.59
OSM-Yelp—Toronto	3	87	91.8	96.6	93.46	96.68	96.77	9.77	5.59

Table A1. *Cont.*

Dataset	N	Min	Q1	Median	Mean	Q3	Max	Range	SD
Rel-HETER	3	100	100	100	100	100	100	0	0
AAS-ECLASS	2	76.7	77.25	77.8	77.8	78.35	78.9	2.2	1.56
Alaska	2	79	79.14	79.29	79.29	79.44	79.58	0.58	0.41
AS	2	63	65.18	67.35	67.35	69.53	71.71	8.71	6.16
Baby Products (Magellan)	2	75.19	77.24	79.29	79.29	81.34	83.39	8.2	5.8
Bikes (Magellan)	2	77.21	82.74	88.28	88.28	93.82	99.35	22.14	15.66
Books (Magellan)	2	76.81	80.65	84.49	84.49	88.33	92.17	15.36	10.86
CiteSeer	2	95	95.57	96.14	96.14	96.71	97.28	2.28	1.61
IMDB-DBPEDIA	2	48.07	54.3	60.54	60.54	66.77	73	24.93	17.63
Markt-Pilot Small	2	87.11	87.32	87.53	87.53	87.74	87.95	0.84	0.59
Markt-Pilot Medium	2	88.7	88.91	89.12	89.12	89.33	89.54	0.84	0.59
Markt-Pilot Large	2	89.9	89.94	89.97	89.97	90.01	90.04	0.14	0.1
Markt-Pilot Full	2	91.17	91.17	91.17	91.17	91.17	91.17	0	0
MCU-MDB	2	73	74.9	76.81	76.81	78.71	80.61	7.61	5.38
Music	2	77.26	82.84	88.43	88.43	94.02	99.6	22.34	15.8
MusicBrainz	2	92.3	93.76	95.22	95.22	96.68	98.14	5.84	4.13
Song	2	78	81.28	84.56	84.56	87.84	91.12	13.12	9.28
SWW-SWG	2	80	82.21	84.42	84.42	86.62	88.83	8.83	6.24
SWW-TOR	2	88	89.36	90.72	90.72	92.08	93.44	5.44	3.85
Synthetic Companies	2	92.22	93.72	95.21	95.21	96.7	98.2	5.98	4.23
WDC Cameras	2	84.76	86.38	88	88	89.63	91.25	6.49	4.59
WDC LSPM Small	2	86.85	87.27	87.69	87.69	88.11	88.53	1.68	1.19
WDC LSPM Medium	2	91.14	91.64	92.14	92.14	92.63	93.13	1.99	1.41
WDC LSPM Large	2	94.48	95.19	95.9	95.9	96.61	97.32	2.84	2.01
WDC LSPM XLarge	2	95.35	95.8	96.24	96.24	96.69	97.14	1.79	1.27
WDC-All-Small	2	91.45	91.9	92.34	92.34	92.79	93.24	1.79	1.27
Zalando	2	73.22	73.42	73.62	73.62	73.81	74.01	0.79	0.56

References

1. Getoor, L.; Machanavajjhala, A. Entity resolution: Theory, practice & open challenges. *Proc. VLDB Endow.* **2012**, *5*, 2018–2019. https://doi.org/10.14778/2367502.2367564.
2. Gupta, A.K.; Kasthurirathne, S.N.; Xu, H.; Li, X.; Ruppert, M.M.; A Harle, C.; Grannis, S.J. A Framework for a Consistent and Reproducible Evaluation of Manual Review for Patient Matching Algorithms. *J. Am. Med. Inform. Assoc.* **2022**, *29*, 2105–2109. https://doi.org/10.1093/jamia/ocac175.
3. Nguyen, L.; Stoové, M.; Boyle, D.; Callander, D.; McManus, H.; Asselin, J.; Guy, R.; Donovan, B.; Hellard, M.; El-Hayek, C. Privacy-Preserving Record Linkage of Deidentified Records Within a Public Health Surveillance System: Evaluation Study. *J. Med. Internet Res.* **2020**, *22*, e16757. https://doi.org/10.2196/16757.
4. Christophides, V.; Efthymiou, V.; Palpanas, T.; Papadakis, G.; Stefanidis, K. An Overview of End-to-End Entity Resolution for Big Data. *ACM Comput. Surv.* **2021**, *53*, 127. https://doi.org/10.1145/3418896.
5. Abedjan, Z.; Chu, X.; Deng, D.; Fernandez, R.C.; Ilyas, I.F.; Ouzzani, M.; Papotti, P.; Stonebraker, M.; Tang, N. Detecting data errors: Where are we and what needs to be done? *Proc. VLDB Endow.* **2016**, *9*, 993–1004. https://doi.org/10.14778/2994509.2994518.
6. Fellegi, I.P.; Sunter, A.B. A Theory for Record Linkage. *J. Am. Stat. Assoc.* **1969**, *64*, 1183–1210. https://doi.org/10.1080/01621459.1969.10501049.
7. Christen, P. *Data Matching: Concepts and Techniques for Record Linkage, Entity Resolution, and Duplicate Detection*; Data-Centric Systems and Applications; Springer: Berlin/Heidelberg, Germany, 2012. https://doi.org/10.1007/978-3-642-31164-2.
8. Stonebraker, M.; Ilyas, I.F. Data Integration : The Current Status and the Way Forward. *IEEE Data Eng. Bull.* **2018**, *41*, 3–9.
9. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention Is All You Need. In *Proceedings of the Advances in Neural Information Processing Systems*; ICLR: Appleton, WI, USA, 2017; Volume 30.
10. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2019; pp. 4171–4186. https://doi.org/10.18653/v1/N19-1423.
11. Brown, T.B.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. In *Proceedings of the Advances in Neural Information Processing Systems*; ICLR: Appleton, WI, USA, 2020.
12. Raffel, C.; Shazeer, N.; Roberts, A.; Lee, K.; Narang, S.; Matena, M.; Zhou, Y.; Li, W.; Liu, P.J. Exploring the limits of transfer learning with a unified text-to-text transformer. *J. Mach. Learn. Res.* **2020**, *21*, 140.
13. Liu, P.; Yuan, W.; Fu, J.; Jiang, Z.; Hayashi, H.; Neubig, G. Pre-train, Prompt, and Predict: A Systematic Survey of Prompting Methods in Natural Language Processing. *ACM Comput. Surv.* **2023**, *55*, 195. https://doi.org/10.1145/3560815.
14. Ding, L.; Lawson, C.; Shapira, P. Rise of Generative Artificial Intelligence in Science. *Scientometrics* **2025**, *130*, 5093–5114. https://doi.org/10.1007/s11192-025-05413-z.
15. Li, Y.; Li, J.; Suhara, Y.; Doan, A.; Tan, W.C. Deep entity matching with pre-trained language models. *Proc. VLDB Endow.* **2020**, *14*, 50–60. https://doi.org/10.14778/3421424.3421431.
16. Akbarian Rastaghi, M.; Kamalloo, E.; Rafiei, D. Probing the Robustness of Pre-trained Language Models for Entity Matching. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*; ACM Digital Library: New York, NY, USA, 2022; pp. 3786–3790. https://doi.org/10.1145/3511808.3557673.
17. Narayan, A.; Chami, I.; Orr, L.; Ré, C. Can Foundation Models Wrangle Your Data? *Proc. VLDB Endow.* **2022**, *16*, 738–746. https://doi.org/10.14778/3574245.3574258.
18. Barlaug, N.; Gulla, J.A. Neural Networks for Entity Matching: A Survey. *ACM Trans. Knowl. Discov. Data* **2021**, *15*, 52. https://doi.org/10.1145/3442200.
19. Awick, J.P.; Marx Gómez, J. Entity Matching in the Era of Language Models: A Structured Literature Review. In *Proceedings of the 2025 25th International Conference on Control Systems and Computer Science (CSCS)*; IEEE: New York, NY, USA, 2025; pp. 368–375. https://doi.org/10.1109/CSCS66924.2025.00061.
20. Tricco, A.C.; Lillie, E.; Zarin, W.; O'Brien, K.K.; Colquhoun, H.; Levac, D.; Moher, D.; Peters, M.D.; Horsley, T.; Weeks, L.; et al. PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation. *Ann. Intern. Med.* **2018**, *169*, 467–473. https://doi.org/10.7326/M18-0850.
21. Ouzzani, M.; Hammady, H.; Fedorowicz, Z.; Elmagarmid, A. Rayyan—A Web and Mobile App for Systematic Reviews. *Syst. Rev.* **2016**, *5*, 210. https://doi.org/10.1186/s13643-016-0384-4.
22. van Rijsbergen, C.J. *Information Retrieval*, 2nd ed.; Butterworths: London, UK, 1979.
23. Teong, K.S.; Soon, L.K.; Su, T.T. Schema-Agnostic Entity Matching using Pre-trained Language Models. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*; ACM Digital Library: New York, NY, USA, 2020; pp. 2241–2244. https://doi.org/10.1145/3340531.3412131.

24. Wu, T.H.; Kao, B.; Wu, Z.; Feng, X.; Song, Q.; Chen, C. MULCE: Multi-level Canonicalization with Embeddings of Open Knowledge Bases. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2020; pp. 315–327. https://doi.org/10.1007/978-3-030-62005-9_23.
25. Brunner, U.; Stockinger, K. Entity Matching with Transformer Architectures—A Step Forward in Data Integration. In Proceedings of EDBT 2020, Copenhagen, Denmark, 30 March–2 April 2020; pp. 463–473. https://doi.org/10.5441/002/edbt.2020.58.
26. Wang, J.; Li, Y.; Hirota, W. Machamp: A Generalized Entity Matching Benchmark. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*; ACM Digital Library: New York, NY, USA, 2021; pp. 4633–4642. https://doi.org/10.1145/3459637.3482008.
27. Peeters, R.; Bizer, C. Dual-objective fine-tuning of BERT for entity matching. *Proc. VLDB Endow.* **2021**, *14*, 1913–1921. https://doi.org/10.14778/3467861.3467878.
28. Chen, R.; Shen, Y.; Zhang, D. GNEM: A Generic One-to-Set Neural Entity Matching Framework. In *Proceedings of the Web Conference 2021*; Association for Computing Machinery: New York, NY, USA, 2021, pp. 1686–1694. https://doi.org/10.1145/3442381.3450119.
29. Jain, A.; Sarawagi, S.; Sen, P. Deep indexed active learning for matching heterogeneous entity representations. *Proc. VLDB Endow.* **2021**, *15*, 31–45. https://doi.org/10.14778/3485450.3485455.
30. Joshi, S.R.; Somani, A.; Roy, S. ReLink: Complete-Link Industrial Record Linkage Over Hybrid Feature Spaces. In *Proceedings of the 2021 IEEE 37th International Conference on Data Engineering (ICDE)*; IEEE: New York, NY, USA, 2021; pp. 2625–2636. https://doi.org/10.1109/icde51399.2021.00293.
31. Li, B.; Miao, Y.; Wang, Y.; Sun, Y.; Wang, W. Improving the Efficiency and Effectiveness for BERT-based Entity Resolution. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 13226–13233. https://doi.org/10.1609/aaai.v35i15.17562.
32. Katakis, N.; Vikatos, P. Entity Linking of Sound Recordings and Compositions with Pre-trained Language Models. In *Proceedings of the 17th International Conference on Web Information Systems and Technologies*; SciTePress: Setúbal, Portugal, 2021; pp. 474–481. https://doi.org/10.5220/0010713900003058.
33. Piché, D.; Zouaq, A.; Gagnon, M.; Font, L. Masked Language Model Entity Matching for Cultural Heritage Data. In *Proceedings of the International Joint Workshop on Semantic Web and Ontology Design for Cultural Heritage (SWODCH 2021)*; CEUR Workshop Proceedings: Bonn, Germany, 2021; Volume 2949, 5p. Available online: https://ceur-ws.org/Vol-2949/paper5.pdf (accessed on 24 August 2026).
34. Balsebre, P.; Yao, D.; Cong, G.; Hai, Z. Geospatial Entity Resolution. In *Proceedings of the ACM Web Conference 2022*; ACM Digital Library: New York, NY, USA, 2022; pp. 3061–3070. https://doi.org/10.1145/3485447.3512026.
35. Paganelli, M.; Buono, F.D.; Baraldi, A.; Guerra, F. Analyzing how BERT performs entity matching. *Proc. VLDB Endow.* **2022**, *15*, 1726–1738. https://doi.org/10.14778/3529337.3529356.
36. Yao, D.; Gu, Y.; Cong, G.; Jin, H.; Lv, X. Entity Resolution with Hierarchical Graph Attention Networks. In *Proceedings of the 2022 International Conference on Management of Data*; ACM Digital Library: New York, NY, USA, 2022; pp. 429–442. https://doi.org/10.1145/3514221.3517872.
37. Peeters, R.; Bizer, C. Cross-Language Learning for Product Matching. In *Proceedings of the Companion Proceedings of the Web Conference 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 236–238. https://doi.org/10.1145/3487553.3524234.
38. Mozdzonek, M.; Wroblewska, A.; Tkachuk, S.; Lukasik, S. Multilingual Transformers for Product Matching – Experiments and a New Benchmark in Polish. In *Proceedings of the 2022 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*; IEEE: New York, NY, USA, 2022; pp. 1–8. https://doi.org/10.1109/fuzz-ieee55066.2022.9882843.
39. Estrada-Valenciano, R.; Muñiz Sánchez, V.; De-la Torre-Gutiérrez, H. An Entity-Matching System Based on Multimodal Data for Two Major E-Commerce Stores in Mexico. *Mathematics* **2022**, *10*, 2564. https://doi.org/10.3390/math10152564.
40. Ye, C.; Jiang, S.; Zhang, H.; Wu, Y.; Shi, J.; Wang, H.; Dai, G. JointMatcher: Numerically-aware entity matching using pre-trained language models with attention concentration. *Knowl.-Based Syst.* **2022**, *251*, 109033. https://doi.org/10.1016/j.knosys.2022.109033.
41. Al Duhayyim, M.; Mesfer Alshahrani, H.; N. Al-Wesabi, F.; Alamgeer, M.; Mustafa Hilal, A.; Ahmed Hamza, M. Relation-Aware Entity Matching Using Sentence-BERT. *Comput. Mater. Contin.* **2022**, *71*, 1581–1595. https://doi.org/10.32604/cmc.2022.020695.
42. Peeters, R.; Bizer, C. Supervised Contrastive Learning for Product Matching. In *Proceedings of the Companion Proceedings of the Web Conference 2022*; Association for Computing Machinery: New York, NY, USA, 2022; pp. 248–251. https://doi.org/10.1145/3487553.3524254.
43. Dou, W.; Shen, D.; Nie, T.; Kou, Y.; Sun, C.; Cui, H.; Yu, G. Empowering Transformer with Hybrid Matching Knowledge for Entity Matching. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2022; pp. 52–67. https://doi.org/10.1007/978-3-031-00129-1_4.

44. Naeim abadi, A.; Nayeem, M.T.; Rafiei, D. Product Entity Matching via Tabular Data. In *Proceedings of the 32nd ACM International Conference on Information and Knowledge Management*; ACM Digital Library: New York, NY, USA, 2023; pp. 4215–4219. https://doi.org/10.1145/3583780.3615172.
45. Mugeni, J.B.; Lynden, S.; Amagasa, T.; Matono, A. AdapterEM: Pre-trained Language Model Adaptation for Generalized Entity Matching using Adapter-tuning. In *Proceedings of the International Database Engineered Applications Symposium Conference*; ACM Digital Library: New York, NY, USA, 2023; pp. 140–147. https://doi.org/10.1145/3589462.3589498.
46. Genossar, B.; Gal, A.; Shraga, R. The Battleship Approach to the Low Resource Entity Matching Problem. *Proc. ACM Manag. Data* **2023**, *1*, 224. https://doi.org/10.1145/3626711.
47. Zeakis, A.; Papadakis, G.; Skoutas, D.; Koubarakis, M. Pre-Trained Embeddings for Entity Resolution: An Experimental Analysis. *Proc. VLDB Endow.* **2023**, *16*, 2225–2238. https://doi.org/10.14778/3598581.3598594.
48. Foua, B.T.; Talburt, J.R.; Xu, X. Large Language Model-Based Representation Learning for Entity Resolution Using Contrastive Learning. In *Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI)*; IEEE: New York, NY, USA, 2023; pp. 15–22. https://doi.org/10.1109/csci62032.2023.00010.
49. Foua, B.T.; Wang, X.; Talburt, J.; Xu, X. Train Once, Match Everywhere: Harnessing Generative Language Models for Entity Matching. In *Proceedings of the 2023 International Conference on Computational Science and Computational Intelligence (CSCI)*; IEEE: New York, NY, USA, 2023; pp. 30–36. https://doi.org/10.1109/csci62032.2023.00012.
50. Ge, C.; Wang, P.; Chen, L.; Liu, X.; Zheng, B.; Gao, Y. CollaborEM: A Self-Supervised Entity Matching Framework Using Multi-Features Collaboration. *IEEE Trans. Knowl. Data Eng.* **2023**, *35*, 12139–12152. https://doi.org/10.1109/tkde.2021.3134806.
51. Bai, H.; Shen, D.; Dou, W.; Nie, T.; Kou, Y. Domain-Generic Pre-Training for Low-Cost Entity Matching via Domain Alignment and Domain Antagonism. In *Proceedings of the 2023 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2023; pp. 1–7. https://doi.org/10.1109/ijcnn54540.2023.10191296.
52. Ghassabi, S.; Behkamal, B.; Milani, M. Leveraging Knowledge Graphs for Matching Heterogeneous Entities and Explanation. In *Proceedings of the 2023 IEEE International Conference on Big Data (BigData)*; IEEE: New York, NY, USA, 2023; pp. 2910–2919. https://doi.org/10.1109/bigdata59044.2023.10386157.
53. Piché, D.; Font, L.; Zouaq, A.; Gagnon, M. Comparing Heuristic Rules and Masked Language Models for Entity Alignment in the Literature Domain. *J. Comput. Cult. Herit.* **2023**, *16*, 62. https://doi.org/10.1145/3606699.
54. Li, Y.; Li, J.; Suhara, Y.; Doan, A.; Tan, W.C. Effective entity matching with transformers. *VLDB J.* **2023**, *32*, 1215–1235. https://doi.org/10.1007/s00778-023-00779-z.
55. Li, S.; Wu, H. Transformer-based Denoising Adversarial Variational Entity Resolution. *J. Intell. Inf. Syst.* **2023**, *61*, 631–650. https://doi.org/10.1007/s10844-022-00773-x.
56. Peeters, R.; Bizer, C. Using ChatGPT for Entity Matching. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2023; pp. 221–230. https://doi.org/10.1007/978-3-031-42941-5_20.
57. Zhang, J.; Sun, H.; Ho, J.C. EMBA: Entity Matching Using Multi-Task Learning of BERT with Attention-over-Attention. In Proceedings of the 27th International Conference on Extending Database Technology (EDBT), Paestum, Italy, 25–28 March 2024. https://doi.org/10.48786/edbt.2024.25.
58. Guermazi, Y.; Sellami, S.; Boucelma, O. GeoRoBERTa: A Transformer-based Approach for Semantic Address Matching. In *Proceedings of the Workshops of the EDBT/ICDT 2023 Joint Conference, DARLI-AP Workshop*; CEUR Workshop Proceedings: Bonn, Germany, 2023; Volume 3379. Available online: https://ceur-ws.org/Vol-3379/DARLI-AP_2023_1.pdf (accessed on 24 August 2026).
59. Xing, X.; Wang, N. Entity Matching Based on Attribute-Aware and Multi-Perspective Similarity Measurement. *J. Inf. Sci. Eng.* **2023**, *39*, 423–438. https://doi.org/10.6688/JISE.202303_39(2).0011.
60. Pyo, J.; Chiang, Y.Y. Leveraging Large Language Models for Generating Labeled Mineral Site Record Linkage Data. In *Proceedings of the 7th ACM SIGSPATIAL International Workshop on AI for Geographic Knowledge Discovery*; ACM Digital Library: New York, NY, USA, 2024; pp. 86–98. https://doi.org/10.1145/3687123.3698298.
61. Dou, W.; Shen, D.; Zhou, X.; Bai, H.; Kou, Y.; Nie, T.; Cui, H.; Yu, G. Enhancing Deep Entity Resolution with Integrated Blocker-Matcher Training: Balancing Consensus and Discrepancy. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*; ACM Digital Library: New York, NY, USA, 2024; pp. 508–518. https://doi.org/10.1145/3627673.3679843.
62. Fan, M.; Han, X.; Fan, J.; Chai, C.; Tang, N.; Li, G.; Du, X. Cost-Effective In-Context Learning for Entity Resolution: A Design Space Exploration. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE)*; IEEE: New York, NY, USA, 2024; pp. 3696–3709. https://doi.org/10.1109/icde60146.2024.00284.
63. Xu, Y.; Li, C.; Yin, Z.; Lu, H.; Han, T.; Yang, Z. Efficient Entity Resolution via Hierarchical Graph Attention and Semantic Blocking. In *Proceedings of the 2024 5th International Conference on Big Data & Artificial Intelligence & Software Engineering (ICBASE)*; IEEE: New York, NY, USA, 2024; pp. 277–281. https://doi.org/10.1109/icbase63199.2024.10762137.

64. Çeliktuğ, M.F.; Kantarcıoğlu, M. Power of Sentence Transformers in Record Linkage. In *Proceedings of the 2024 IEEE International Conference on Big Data (BigData)*; IEEE: New York, NY, USA, 2024; pp. 6944–6955. https://doi.org/10.1109/bigdata62323.2024.10825999.
65. Yuedanni. Adaptive Target-Consistency Entity Matching Algorithm Based on Semi-Supervised Learning. In *Proceedings of the 2024 10th International Conference on Big Data and Information Analytics (BigDIA)*; IEEE: New York, NY, USA, 2024; pp. 31–37. https://doi.org/10.1109/bigdia63733.2024.10808744.
66. Rettenmeier, T.; Jesser, A. Data Augmentation for Entity Resolution: A Comparative Evaluation. In *Proceedings of the 2024 IEEE 3rd International Conference on Computing and Machine Intelligence (ICMI)*; IEEE: New York, NY, USA, 2024; pp. 1–5. https://doi.org/10.1109/icmi60790.2024.10585627.
67. Cao, H.; Du, S.; Hu, J.; Yang, Y.; Horng, S.J.; Li, T. Graph Deep Active Learning Framework for Data Deduplication. *Big Data Min. Anal.* **2024**, *7*, 753–764. https://doi.org/10.26599/bdma.2023.9020040.
68. Zeng, X.; Wang, P.; Mao, Y.; Chen, L.; Liu, X.; Gao, Y. MultiEM: Efficient and Effective Unsupervised Multi-Table Entity Matching. In *Proceedings of the 2024 IEEE 40th International Conference on Data Engineering (ICDE)*; IEEE: New York, NY, USA, 2024; pp. 3421–3434. https://doi.org/10.1109/icde60146.2024.00264.
69. Song, H.; Liu, M.; Zhang, S.; Han, Q. Dual-Module Feature Alignment Domain Adversarial Model for Entity Resolution. In *Proceedings of the 2024 11th International Conference on Behavioural and Social Computing (BESC)*; IEEE: New York, NY, USA, 2024; pp. 1–8. https://doi.org/10.1109/besc64747.2024.10780643.
70. Lian, Z.; Tie-zheng, N.; Shen, D.; Kou, Y. LRER: A Low-Resource Entity Resolution Framework with Hybrid Information. In *Proceedings of the 2024 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2024; pp. 1–8. https://doi.org/10.1109/ijcnn60899.2024.10651166.
71. Geng, M. Enhancing entity resolution with multichannel BERT: A comprehensive approach. In *Proceedings of the Third International Conference on Algorithms, Microchips, and Network Applications (AMNA 2024)*; SPIE: Bellingham, WA, USA, 2024, p. 17. https://doi.org/10.1117/12.3031934.
72. Paganelli, M.; Tiano, D.; Del Buono, F.; Baraldi, A.; Benassi, R.; Guiduzzi, G.; Guerra, F. How Transformers Are Revolutionizing Entity Matching. In *Proceedings of the 32nd Italian Symposium on Advanced Database Systems (SEBD 2024)*; CEUR Workshop Proceedings: Bonn, Germany, 2024; Volume 3741, 55p. Available online: https://ceur-ws.org/Vol-3741/paper55.pdf (accessed on 24 August 2026).
73. Nuntachit, N.; Sugannasil, P. Can ChatGPT Outperform Other Language Models? An Experiment on Using ChatGPT for Entity Matching Versus Other Language Models. In *Lecture Notes on Data Engineering and Communications Technologies*; Springer: Cham, Switzerland, 2024; pp. 14–26. https://doi.org/10.1007/978-3-031-46970-1_2.
74. Ding, H.; Dai, C.; Wu, Y.; Ma, W.; Zhou, H. SETEM: Self-ensemble training with Pre-trained Language Models for Entity Matching. *Knowl.-Based Syst.* **2024**, *293*, 111708. https://doi.org/10.1016/j.knosys.2024.111708.
75. Pollack, J.; Köpcke, H.; Rahm, E. Intermediate Fusion for Multimodal Product Matching. In *Proceedings of the 35th GI-Workshop on Foundations of Databases (GvDB 2024)*; CEUR Workshop Proceedings: Bonn, Germany, 2024. Available online: https://dbs.uni-leipzig.de/files/research/publications/2024-5/pdf/gvdb2024_cameraReady.pdf (accessed on 24 August 2026).
76. Low, J.F.; Fung, B.C.; Xiong, P. Better entity matching with transformers through ensembles. *Knowl.-Based Syst.* **2024**, *293*, 111678. https://doi.org/10.1016/j.knosys.2024.111678.
77. Pardo, F.D.M.; Lehmann, C.; Gehrig, D.; Nagy, A.; Nicoli, S.; Misheva, B.H.; Braschler, M.; Stockinger, K. GraLMatch: Matching Groups of Entities with Graphs and Language Models. In Proceedings of the 28th International Conference on Extending Database Technology (EDBT), Barcelona, Spain, 25–28 March 2025. https://doi.org/10.48786/edbt.2025.01.
78. Nananukul, N.; Sisaengsuwanchai, K.; Kejriwal, M. Cost-efficient prompt engineering for unsupervised entity resolution in the product matching domain. *Discov. Artif. Intell.* **2024**, *4*, 56. https://doi.org/10.1007/s44163-024-00159-8.
79. Zhu, L.; Liu, H.; Song, X.; Wei, Y.; Wang, Y. Entity Resolution Based on Pre-trained Language Models with Two Attentions. In *Lecture Notes in Computer Science*; Springer: Singapore, 2024; pp. 433–448. https://doi.org/10.1007/978-981-97-2387-4_29.
80. Tang, J.; Dou, W.; Shen, D.; Nie, T.; Kou, Y. Towards Long-Text Entity Resolution with Chain-of-Thought Knowledge Augmentation from Large Language Models. In *Lecture Notes in Computer Science*; Springer: Singapore, 2024; pp. 322–336. https://doi.org/10.1007/978-981-97-5569-1_20.
81. Wadhwa, S.; Krishnan, A.; Wang, R.; Wallace, B.C.; Kong, L. Learning from Natural Language Explanations for Generalizable Entity Matching. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024; pp. 6114–6129. https://doi.org/10.18653/v1/2024.emnlp-main.352.
82. Paul, A.; Maheshwary, S.; Sohoney, S. Accurate Customer Address Matching via Weak Supervision for Geocode Learning. In *Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems*; ACM Digital Library: New York, NY, USA, 2024; pp. 454–464. https://doi.org/10.1145/3678717.3691277.
83. Jabrane, M.; Toulaoui, A.; Hafidi, I. Enhancing Semantic Web Entity Matching Process Using Transformer Neural Networks and Pre-Trained Language Models. *Comput. Informatics* **2024**, *43*, 1397–1415. https://doi.org/10.31577/cai_2024_6_1397.

84. Arora, A.; Dell, M. LinkTransformer: A Unified Package for Record Linkage with Transformer Language Models. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 3: System Demonstrations)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024; pp. 221–231. https://doi.org/10.18653/v1/2024.acl-demos.21.
85. Wang, Y.; Zhou, L.; Wang, Y.; Peng, Z. Leveraging Pretrained Language Models for Enhanced Entity Matching: A Comprehensive Study of Fine-Tuning and Prompt Learning Paradigms. *Int. J. Intell. Syst.* **2024**, *2024*, 1–14. https://doi.org/10.1155/2024/1941221.
86. He, L.; Li, H.; Zhang, R. A Semantic-Spatial Aware Data Conflation Approach for Place Knowledge Graphs. *ISPRS Int. J. Geo-Inf.* **2024**, *13*, 106. https://doi.org/10.3390/ijgi13040106.
87. Peeters, R.; Steiner, A.; Bizer, C. Entity Matching using Large Language Models. In Proceedings of the 28th International Conference on Extending Database Technology (EDBT), Barcelona, Spain, 25–28 March 2025. https://doi.org/10.48786/edbt.2025.42.
88. Zhang, Z.; Balsebre, P.; Luo, S.; Hai, Z.; Huang, J. StructAM: Enhancing Address Matching through Semantic Understanding of Structure-aware Information. In *Proceedings of the Language Resources and Evaluation Conference*; European Language Resources Association (ELRA): Paris, France; ICCL: Brussels, Belgium, 2024; pp. 15350–15361. https://doi.org/10.63317/56c3joqfyz4a.
89. Wang, R.; Tao, Y.; Krishnan, A.; Kong, L.; Liu, X.; Deng, Y.; Yang, Y.; Johnson, H.; Borthwick, A.; Gupta, S.; et al. BPID: A Benchmark for Personal Identity Deduplication. In *Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing: Industry Track*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2024; pp. 538–546. https://doi.org/10.18653/v1/2024.emnlp-industry.40.
90. Alves, A.L.F.; Baptista, C.d.S.; Barbosa, L.; Araujo, C.B.M. Cross-Lingual Learning Strategies for Improving Product Matching Quality. In *Proceedings of the 39th ACM/SIGAPP Symposium on Applied Computing*; ACM Digital Library: New York, NY, USA, 2024; pp. 313–320. https://doi.org/10.1145/3605098.3636001.
91. Fu, J.; Tang, H.; Khan, A.; Mehrotra, S.; Ke, X.; Gao, Y. In-context Clustering-based Entity Resolution with Large Language Models: A Design Space Exploration. *Proc. ACM Manag. Data* **2025**, *3*, 252. https://doi.org/10.1145/3749170.
92. Steiner, A.; Peeters, R.; Bizer, C. Fine-Tuning Large Language Models for Entity Matching. In *Proceedings of the 2025 IEEE 41st International Conference on Data Engineering Workshops (ICDEW)*; IEEE: New York, NY, USA, 2025; pp. 9–17. https://doi.org/10.1109/icdew67478.2025.00006.
93. Saengsiripaiboon, S.; Pacharawongsakda, E.; Jitkongchuen, D. Enhancing Efficiency in Entity Resolution Strategies for Batch Prompting. In *Proceedings of the 2025 6th International Conference on Big Data Analytics and Practices (IBDAP)*; IEEE: New York, NY, USA, 2025; pp. 31–35. https://doi.org/10.1109/ibdap65587.2025.11145850.
94. Bega, M.; Kuhlenkötter, B. Utilizing the Mixture-of-Agents Approach for Entity Resolution and Data Landscape Homogenization in Manufacturing Domains. In *Proceedings of the 2025 IEEE 8th International Conference on Industrial Cyber-Physical Systems (ICPS)*; IEEE: New York, NY, USA, 2025; pp. 1–4. https://doi.org/10.1109/icps65515.2025.11087844.
95. Arford, D.; Lin, F.; Li, W.; Hu, J.; Chen, H. Assessing the De-Anonymization Risk of Social Media Users: A BERT-Based Entity Resolution Approach. In *Proceedings of the 2025 IEEE International Conference on Intelligence and Security Informatics (ISI)*; IEEE: New York, NY, USA, 2025; pp. 15–20. https://doi.org/10.1109/isi65680.2025.11201172.
96. Karapiperis, D.; Feretzakis, G.; Verykios, V.S. An Experimental Evaluation of Pre-Trained Models for Efficient and Accurate Record Linkage. In *Proceedings of the 2025 16th International Conference on Information, Intelligence, Systems & Applications (IISA)*; IEEE: New York, NY, USA, 2025; pp. 1–6. https://doi.org/10.1109/iisa66859.2025.11311205.
97. Song, H.; Zhang, S.; Zhao, Y.; Han, Q. Unveiling Hidden Gems: Enhancing Entity Resolution with a Data Perspective. In *Proceedings of the 2025 International Joint Conference on Neural Networks (IJCNN)*; IEEE: New York, NY, USA, 2025; pp. 1–8. https://doi.org/10.1109/ijcnn64981.2025.11229070.
98. Liu, C.; Rong, X. Automated Graph Attention Network for Heterogeneous Entity Resolution. In *Proceedings of the ICASSP 2025—2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; IEEE: New York, NY, USA, 2025; pp. 1–5. https://doi.org/10.1109/icassp49660.2025.10890537.
99. Sharifi, M.; Ahmadzadeh, D. Transformer-Gather, Fuzzy-Reconsider: A Scalable Hybrid Framework for Entity Resolution. In *Proceedings of the 2025 15th International Conference on Computer and Knowledge Engineering (ICCKE)*; IEEE: New York, NY, USA, 2025; pp. 1–6. https://doi.org/10.1109/iccke68588.2025.11273830.
100. Seo, W.; Shin, T.; An, H.; Kim, D.; Lee, S. Question-to-Knowledge (Q2K): Multi-Agent Generation of Inspectable Facts for Product Mapping. In *Proceedings of the 2025 IEEE International Conference on Big Data (BigData)*; IEEE: New York, NY, USA, 2025; pp. 2646–2653. https://doi.org/10.1109/bigdata66926.2025.11401468.
101. Haruki, Y.; Ishikura, S.; Demachi, K.; Hayashi, T. Contextual Graph Embeddings: Accounting for Data Characteristics in Heterogeneous Data Integration. In *Proceedings of the 2025 IEEE International Conference on Big Data (BigData)*; IEEE: New York, NY, USA, 2025; pp. 2843–2852. https://doi.org/10.1109/bigdata66926.2025.11400991.

102. Yuan, Q.; Yuan, Y.; Wen, Z.; Chen, C.; Wang, G. CrossEM: A Prompt Tuning Framework for Cross-Modal Entity Matching. In *Proceedings of the 2025 IEEE 41st International Conference on Data Engineering (ICDE)*; IEEE: New York, NY, USA, 2025; pp. 627–640. https://doi.org/10.1109/icde65448.2025.00053.
103. Kulunk, A.; Taskin, B.; Eseoglu, M.F.; Sahin, H.B. Optimizing Product Deduplication in E-Commerce with Multimodal Embeddings. In *Proceedings of the 2025 IEEE International Conference on Big Data (BigData)*; IEEE: New York, NY, USA, 2025; pp. 2577–2584. https://doi.org/10.1109/bigdata66926.2025.11401681.
104. Ahmed, M.O.S. A Late-Fusion Multimodal AI Framework for Privacy-Preserving Deduplication in National Healthcare Data Environments. In *Proceedings of the 2025 IEEE International Conference on Future Machine Learning and Data Science (FMLDS)*; IEEE: New York, NY, USA, 2025; pp. 74–80. https://doi.org/10.1109/fmlds67896.2025.00021.
105. Pardo, F.D.M.; Hadji Misheva, B.; Braschler, M.; Stockinger, K. TransClean: Finding False Positives in Multi-Source Entity Matching Under Real-World Conditions via Transitive Consistency. *IEEE Access* **2025**, *13*, 195856–195870. https://doi.org/10.1109/access.2025.3632400.
106. Li, X.; Fan, S.; Yao, J.; Sun, H. Contextual semantics graph attention network model for entity resolution. *Sci. Rep.* **2025**, *15*, 27093. https://doi.org/10.1038/s41598-025-11932-9.
107. Shi, D.; Meyer, O.; Oberle, M.; Bauernhansl, T. Dual data mapping with fine-tuned large language models and asset administration shells toward interoperable knowledge representation. *Robot. Comput.-Integr. Manuf.* **2025**, *91*, 102837. https://doi.org/10.1016/j.rcim.2024.102837.
108. Yan, M.; Wang, Y.; Jiang, X.; Zhou, H.; Li, J. Towards uncertainty-calibrated structural data enrichment with large language model for few-shot entity resolution. *Front. Comput. Sci.* **2025**, *19*, 1911376. https://doi.org/10.1007/s11704-025-41143-4.
109. Yin, H.; Kim, J.; Mathur, P.; Sarker, K.; Bansal, V. How to Talk to Language Models: Serialization Strategies for Structured Entity Matching. In *Proceedings of the Findings of the Association for Computational Linguistics: NAACL 2025*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 7836–7850. https://doi.org/10.18653/v1/2025.findings-naacl.437.
110. Arvanitis-Kasinikos, I.; Papadakis, G. Entity Matching with 7B LLMs: A Study on Prompting Strategies and Hardware Limitations. In *Proceedings of the International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP 2025)*; CEUR Workshop Proceedings: Bonn, Germany, 2025; Volume 3931, pp. 31–38. Available online: https://ceur-ws.org/Vol-3931/paper4.pdf (accessed on 24 August 2026).
111. Santana, D.; Lima, P.; Ribeiro, L. EM-Join: Efficient Entity Matching Using Embedding-Based Similarity Join. In *Proceedings of the 27th International Conference on Enterprise Information Systems*; SciTePress: Setúbal, Portugal, 2025; pp. 402–409. https://doi.org/10.5220/0013483700003929.
112. Alves, A.; Baptista, C.; Diniz, J.; Mendes, F.; Cunha, M. An Approach for Product Record Linkage Using Cross-Lingual Learning and Large Language Models. In *Proceedings of the 27th International Conference on Enterprise Information Systems*; SciTePress: Setúbal, Portugal, 2025; pp. 63–74. https://doi.org/10.5220/0013285000003929.
113. Zeakis, A.; Papadakis, G.; Skoutas, D.; Koubarakis, M. AvengER: Ensembling and Fine-Tuning LLMs for SELECT Prompts in Entity Resolution. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2025; pp. 301–320. https://doi.org/10.1007/978-3-031-94575-5_17.
114. Shah, N.; Patiyara, S.; Basak, J.; Sahni, S.; Mathur, A.; Park, K.; Rajasekaran, S. Efficient Record Linkage in the Age of Large Language Models: The Critical Role of Blocking. *Algorithms* **2025**, *18*, 723. https://doi.org/10.3390/a18110723.
115. Ruan, Q.; Shi, D.; Bauernhansl, T. Fine-tuning large language models with contrastive margin ranking loss for selective entity matching in product data integration. *Adv. Eng. Inform.* **2025**, *67*, 103538. https://doi.org/10.1016/j.aei.2025.103538.
116. Ornstein, J.T. Probabilistic Record Linkage Using Pretrained Text Embeddings. *Political Anal.* **2026**, *34*, 205–216. https://doi.org/10.1017/pan.2025.10016.
117. Wang, S.; Miao, S.; Kwashie, S.; Bewong, M.; Hu, J.; Feng, Z. LLM-Enhanced Entity Resolution Using Graph Differential Dependencies. In *Lecture Notes in Computer Science*; Springer: Singapore, 2025; pp. 125–136. https://doi.org/10.1007/978-981-96-9921-6_11.
118. Kardos, P.; Vass, M.; Krész, M.; Farkas, R. DogMa results for OAEI 2025. In *Proceedings of the Ontology Matching Workshop (OM 2025), Co-Located with ISWC 2025*; CEUR Workshop Proceedings: Bonn, Germany, 2025; Volume 4144, pp. 211–215. Available online: https://ceur-ws.org/Vol-4144/om2025-oaei-paper12.pdf (accessed on 24 August 2026).
119. Wang, T.; Chen, X.; Lin, H.; Chen, X.; Han, X.; Sun, L.; Wang, H.; Zeng, Z. Match, Compare, or Select? An Investigation of Large Language Models for Entity Matching. In *Proceedings of the 31st International Conference on Computational Linguistics (COLING 2025)*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 96–109. Available online: https://aclanthology.org/2025.coling-main.8/ (accessed on 24 August 2026).
120. Rettenmeier, T.; Jesser, A. Graph-Based Data Augmentation and Label Noise Identification for Entity Resolution. In *Communications in Computer and Information Science*; Springer: Singapore, 2025; pp. 47–61. https://doi.org/10.1007/978-981-96-7005-5_4.

121. Wang, Y.; Yan, M.; Wang, W. PUER: Boosting Few-shot Positive-Unlabeled Entity Resolution with Reinforcement Learning. In *Proceedings of the Findings of the Association for Computational Linguistics: EMNLP 2025*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 24567–24579. https://doi.org/10.18653/v1/2025.findings-emnlp.1336.
122. Zhu, H.; Li, Z.; Jin, J. GER-LLM: Efficient and Effective Geospatial Entity Resolution with Large Language Model. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing*; Association for Computational Linguistics: Stroudsburg, PA, USA, 2025; pp. 23272–23288. https://doi.org/10.18653/v1/2025.emnlp-main.1186.
123. Mugeni, J.B.; Lynden, S.; Amagasa, T.; Matono, A. AssistEM: Domain Instruction Tuning for Enhanced Entity Matching. In *Lecture Notes in Computer Science*; Springer: Singapore, 2025; pp. 115–127. https://doi.org/10.1007/978-981-96-8186-0_10.
124. Yamamoto, V.E.; Takeda, H. Using LLM to Improve Knowledge Graph Entity Matching. In *Proceedings of the ISWC 2025 Companion Volume*; CEUR Workshop Proceedings: Bonn, Germany, 2025; Volume 4085, 32p. Available online: https://ceur-ws.org/Vol-4085/paper32.pdf (accessed on 24 August 2026).
125. Sun, Z.; Liu, P.; Zhai, L.; Zhang, Z. A Geometric Attribute Collaborative Method in Multi-Scale Polygonal Entity Matching Scenario: Integrating Sentence-BERT and Three-Branch Attention Network. *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 435. https://doi.org/10.3390/ijgi14110435.
126. Ji, Y.; Xie, W.; Zhang, J.; Wang, C.; Guo, N.; Shi, L.; Zhang, Y. Unaligned Message-Passing and Contextualized-Pretraining for Robust Geo-Entity Resolution. *Proc. AAAI Conf. Artif. Intell.* **2025**, *39*, 11852–11860. https://doi.org/10.1609/aaai.v39i11.33290.
127. Zeakis, A.; Papadakis, G.; Skoutas, D.; Koubarakis, M. An in-depth analysis of pre-trained embeddings for entity resolution. *VLDB J.* **2025**, *34*, 5. https://doi.org/10.1007/s00778-024-00879-4.
128. O'Reilly-Morgan, D.; Tragos, E.; Duriakova, E.; Du, H.; Hurley, N.; Lawlor, A. Entity Matching with Large Language Models as Weak and Strong Labellers. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2025; pp. 58–67. https://doi.org/10.1007/978-3-031-70421-5_6.
129. Nananukul, N.; Kekriwal, M. Balancing Efficiency and Quality in LLM-Based Entity Resolution on Structured Data. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2025; pp. 278–293. https://doi.org/10.1007/978-3-031-78548-1_21.
130. Olar, A. On the Asymmetrical Nature of Entity Matching Using Pre-Trained Transformers. In *Proceedings of the 17th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*; SciTePress: Setúbal, Portugal, 2025; pp. 208–215. https://doi.org/10.5220/0013672900004000.
131. Shi, H.; Liu, X.; Lv, F.; Xue, H.; Hu, J.; Du, S.; Li, T. A pre-trained data deduplication model based on active learning. *Expert Syst. Appl.* **2025**, *292*, 128628. https://doi.org/10.1016/j.eswa.2025.128628.
132. Zhang, Z.; Groth, P.; Calixto, I.; Schelter, S. A Deep Dive Into Cross-Dataset Entity Matching with Large and Small Language Models. In Proceedings of the 28th International Conference on Extending Database Technology (EDBT), Barcelona, Spain, 25–28 March 2025. https://doi.org/10.48786/edbt.2025.75.
133. Du, N. A multi-layer representation mixing method based on BERT for entity resolution. In *Proceedings of the Fourth International Conference on Algorithms, Microchips, and Network Applications (AMNA 2025)*; SPIE: Bellingham, WA, USA, 2025, p. 9. https://doi.org/10.1117/12.3068420.
134. Zhang, Z.; Zeng, W.; Tang, J.; Huang, H.; Zhao, X. Active in-context learning for cross-domain entity resolution. *Inf. Fusion* **2025**, *117*, 102816. https://doi.org/10.1016/j.inffus.2024.102816.
135. Althaf, A.M.; Morshed, M.S.; Kabir, M.R.; Milanova, M.; Talburt, J. Semantic Entity Resolution on Synthetic Datasets: A Transformer-Centric Approach. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2025; pp. 107–116. https://doi.org/10.1007/978-3-031-88220-3_7.
136. Althaf, A.M.; Mohammed, M.A.; Milanova, M.; Talburt, J.; Cakmak, M.C. Multi-Agent RAG Framework for Entity Resolution: Advancing Beyond Single-LLM Approaches with Specialized Agent Coordination. *Computers* **2025**, *14*, 525. https://doi.org/10.3390/computers14120525.
137. Mosa, M.A. Synergizing structure and semantics: A knowledge graph-transformer framework for narrator disambiguation in hadith networks. *Digit. Scholarsh. Humanit.* **2025**, *40*, 1085–1100. https://doi.org/10.1093/llc/fqaf088.
138. Zhang, J.; Fang, J.; Zhang, C.; Zhang, W.; Ren, H.; Xu, L. Geographic Named Entity Matching and Evaluation Recommendation Using Multi-Objective Tasks: A Study Integrating a Large Language Model (LLM) and Retrieval-Augmented Generation (RAG). *ISPRS Int. J. Geo-Inf.* **2025**, *14*, 95. https://doi.org/10.3390/ijgi14030095.
139. Barreto, C.A.S.; Junior, A.S.B.; Melo, L.D.G.; Santos, M.D.R.; Carvalho, G.R.; Filho, I.M.B.; Neto, J.G.R.; Malaquias, R.S.; Gurgel, A.M.; Lima, J.M.M. Development of a Strategy for Duplication Search Based on Multiple Hierarchically Organized Approaches. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2025; pp. 368–385. https://doi.org/10.1007/978-3-031-96997-3_23.
140. Kabir, M.R.; Althaf, A.M.; Morshed, M.S.; Milanova, M.; Talburt, J. Entity Resolution Using Transformers for Synthetic Datasets. In *Smart Innovation, Systems and Technologies*; Springer: Singapore, 2025; pp. 113–123. https://doi.org/10.1007/978-981-96-2482-9_8.
141. Chen, X.; Shi, Y.; Xiao, X. Generalized Entity Matching with Adaptivity via Large Language Models. *Proc. ACM Manag. Data* **2026**, *4*, 189. https://doi.org/10.1145/3802066.

142. Pulsone, N.; Shraga, R.; Goren, G. BEACON: Budget-Aware Entity Matching Across Domains. *Proc. ACM Manag. Data* **2026**, *4*, 144. https://doi.org/10.1145/3802021.
143. Zhang, S.; Zeng, W.; Zhang, Z.; Hou, W.; Xiao, W.; Zhao, X. CoMCo: Consistency-Aware Multi-Agent Coordination for Zero-Shot Cross-Modal Entity Matching. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*; ACM Digital Library: New York, NY, USA, 2026; pp. 2398–2409. https://doi.org/10.1145/3805712.3809543.
144. Zhang, Z.; Zeng, W.; Tang, J.; Zhao, X. Unified Entity Matching under Scarce Supervision via Meta-Rule Induction and Retrieval. In *Proceedings of the 49th International ACM SIGIR Conference on Research and Development in Information Retrieval*; ACM Digital Library: New York, NY, USA, 2026; pp. 2386–2397. https://doi.org/10.1145/3805712.3809619.
145. Wang, H.; Yang, R.; Zheng, H.; Ke, X. Adaptive Graph Refinement and Label Propagation with LLMs for Cost-Effective Entity Resolution. In *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2*; ACM Digital Library: New York, NY, USA, 2026; pp. 4906–4915. https://doi.org/10.1145/3770855.3817856.
146. Akashi, K.; Ogawa, M.; Tayama, K. Cost-Effective Entity Matching with Large Language Models and Diversity-Aware Example Selection. *IEEE Access* **2026**, *14*, 105851–105862. https://doi.org/10.1109/access.2026.3708863.
147. He, S.; Sen, L.; Tang, L.; Jiang, X. Enhancing Nuclear Power Plant Data Inventory Generation Through Large Language Models: A Framework for Efficient Entity Matching. In *Proceedings of the 2026 5th International Conference on Electronics Technology and Artificial Intelligence (ETAI)*; IEEE: New York, NY, USA, 2026; pp. 67–71. https://doi.org/10.1109/etai68332.2026.11485364.
148. Xu, Y.; Zhai, H.; Song, C.; Shen, H. IDP-EM: Instance-Aware Dynamic Prompt Tuning With Pseudo-Label-Enhanced Contrastive Learning for Low-Resource Entity Matching. *IEEE Access* **2026**, *14*, 116713–116728. https://doi.org/10.1109/access.2026.3718155.
149. Song, J.; Cao, Y.; Wang, Y.; Mu, C.; An, Y. CrossFuse: A Cross-Attentive Intermediate Fusion Framework for Multimodal Entity Matching. In *Proceedings of the 2026 11th International Conference on Computer and Communication Systems (ICCCS)*; IEEE: New York, NY, USA, 2026; pp. 1125–1131. https://doi.org/10.1109/icccs69761.2026.11613502.
150. Şahin, U.B.; Türkmen, H.; Çiftlikçi, M.S.; Arık, İ.U.; Uyan, U.; Kalaycı, T.A.; Çakar, T. Multimodal Product Unification for Large-Scale E-commerce Catalogs. In *Proceedings of the 2026 8th International Congress on Human-Computer Interaction, Optimization and Robotic Applications (ICHORA)*; IEEE: New York, NY, USA, 2026; pp. 1–6. https://doi.org/10.1109/ichora69329.2026.11537143.
151. Bryan, K.; Tsapara, I. Entity Resolution for Aircraft Type Matching on Heterogeneous Aviation Data Sources. In *Proceedings of the 2026 IEEE Aerospace Conference*; IEEE: New York, NY, USA, 2026; pp. 1–10. https://doi.org/10.1109/aero66936.2026.11519903.
152. Cao, C.; Pillai, J.; Daraei, S.; Ghadermarzi, S. Linking patient records at scale with a hybrid approach combining contrastive learning and deterministic rules. *Biol. Methods Protoc.* **2026**, *11*, bpag009. https://doi.org/10.1093/biomethods/bpag009.
153. Mohammed, M.A.; Talburt, J.; Althaf, A.M.; Milanova, M. Multi-LLM Record Linkage: A Comparative Analysis Framework for Co-Residence Pattern Discovery. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2027; pp. 436–451. https://doi.org/10.1007/978-3-032-22196-4_32.
154. Liu, X.; Li, X.; Yao, J.; Liu, Y.; Fan, Q.; Sun, H.; Dong, C. CGEM: A cognitive-guided network for human-aligned entity matching. *Neurocomputing* **2026**, *675*, 132950. https://doi.org/10.1016/j.neucom.2026.132950.
155. Thivaios, G.; Zervas, P.; Giotopoulos, K.; Tzimas, G. On the Task of Job Posting Deduplication Using Embedding-Based Filtering and LLM Validation. *Information* **2026**, *17*, 233. https://doi.org/10.3390/info17030233.
156. Okayama, K.; Ito, H.; Morishima, A. Learning from unknown-unknowns: Inconsistency-driven sampling for improving LLM entity matching. *Inf. Res. Int. Electron. J.* **2026**, *31*, 118–135. https://doi.org/10.47989/ir31iconf64127.
157. Eseoglu, M.F.; Kulunk, A.; Taskin, B.; Sahin, H.B. Enhancing E-Commerce Product Matching with a Hybrid AI Approach: A Three-Stage System with LLM-based Verification. In *Proceedings of the 2025 8th Artificial Intelligence and Cloud Computing Conference*; ACM Digital Library: New York, NY, USA, 2025; pp. 567–574. https://doi.org/10.1145/3789982.3790065.
158. Tian, Y.; Wang, N.; Zhou, A. CrossER: A robust and adaptable generalized entity resolution framework for diverse and heterogeneous datasets. *Inf. Syst.* **2026**, *135*, 102609. https://doi.org/10.1016/j.is.2025.102609.
159. Karapiperis, D.; Akritidis, L.; Bozanis, P. The impact of fine-tuning on entity resolution: An experimental evaluation. *Knowl.-Based Syst.* **2026**, *338*, 115427. https://doi.org/10.1016/j.knosys.2026.115427.
160. Yang, D.; Xiong, M.; Shen, D.; Nie, T.; Kou, Y. KIKA: Dual-stage domain knowledge injection–adaptation for multi-domain entity matching with contextual prompt enhancement. *J. Supercomput.* **2026**, *82*, 57. https://doi.org/10.1007/s11227-025-08160-3.
161. Xiong, M.; Ma, H.; Shen, D.; Nie, T.; Kou, Y. FSEM: Few-Shot Entity Matching Using Multi-loss Adversarial Training with Multi-attention Masking. In *Lecture Notes in Computer Science*; Springer: Singapore, 2026; pp. 646–662. https://doi.org/10.1007/978-981-92-0366-6_39.
162. Mohammed, M.A.; Mandalawi, S.A.; Maclean, H.; Talburt, J.R. Multilingual Customer Record Linkage: A Novel Approach Using LLMs for Cross-Lingual Entity Resolution. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2027; pp. 391–405. https://doi.org/10.1007/978-3-032-22196-4_29.
163. Wijegunarathna, K.; Stock, K.; Jones, C.B. Omni Geometry Representation Learning Versus Large Language Models for Geospatial Entity Resolution. *Trans. GIS* **2026**, *30*, e70199. https://doi.org/10.1111/tgis.70199.

164. Mekki, N.; Berrabah, D.; Malki, A. DeepIDDFS: Efficient entity resolution with hybrid embeddings. *Knowl. Inf. Syst.* **2026**, *68*, 229. https://doi.org/10.1007/s10115-026-02847-6.
165. Dou, W.; Shen, D.; Zhou, X.; Kou, Y.; Nie, T.; Cui, H.; Yu, G. Weakly-supervised entity matching via LLM-guided data augmentation and knowledge transfer. *Knowl.-Based Syst.* **2026**, *335*, 115238. https://doi.org/10.1016/j.knosys.2025.115238.
166. Neuhof, F.; Fisichella, M.; Papadakis, G. End-to-End Deep Entity Resolution without Labelled Instances. In *Proceedings of the International Workshop on Design, Optimization, Languages and Analytical Processing of Big Data (DOLAP 2026)*; CEUR Workshop Proceedings: Bonn, Germany, 2026; Volume 4186, pp. 100–106. Available online: https://ceur-ws.org/Vol-4186/short2.pdf (accessed on 24 August 2026).
167. Tang, Y.; Su, T.; Zhang, W.; Guo, X.; Liu, T. Unlocking the Power of Large Language Models for Multi-table Entity Matching. In *Lecture Notes in Computer Science*; Springer: Singapore, 2026; pp. 208–220. https://doi.org/10.1007/978-981-95-3352-7_17.
168. Xu, Y.; Yang, Z.; Liang, S.; Liu, Y.; Xie, C.; Zhai, H.; Shi, X. MIEM-CA: A Multigranularity Information-Enhanced Entity Matching Method Based on Collaborative Agents. *Int. J. Intell. Syst.* **2026**, *2026*, 8100559. https://doi.org/10.1155/int/8100559.
169. Gomes Ribeiro, A.E.; Ferreira Rodrigues Moreira, L. Profile Identification on Social Media Using Artificial Intelligence Techniques for Textual Similarity. *Rev. De Informática Teórica E Apl.* **2026**, *33*, 59–70. https://doi.org/10.22456/2175-2745.150441.
170. Gvasalia, S.; Pelucchi, M.; Perego, S. Hybrid Ontology Matching for Company Name Alignment: Combining Text Matching, String Similarity, SBERT, and Siamese Networks in Italian and German Job Market Data. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2026; pp. 193–205. https://doi.org/10.1007/978-3-032-05607-8_19.
171. Ather, H. LLM-assisted record linkage: A framework for official statistics. *Stat. J. IAOS* **2026**, *42*, 211–217. https://doi.org/10.1177/18747655261422068.
172. Arvanitis-Kasinikos, I.; Stetsikas, L.; Papadakis, G.; Koubarakis, M. ChatMatcher: End-to-end entity resolution with 7B LLM-based matching. *Inf. Syst.* **2026**, *142*, 102784. https://doi.org/10.1016/j.is.2026.102784.
173. Huang, D.; Pai, P.; Wang, Z. Entity Matching with LLMs at Scale: Accuracy, Cost, and Reasoning Effects. In *Lecture Notes in Computer Science*; Springer: Singapore, 2027; pp. 112–124. https://doi.org/10.1007/978-981-92-2014-4_10.
174. Mugeni, J.B.; Lynden, S.; Amagasa, T.; Matono, A. Optimizing sample selection for large language model-based entity matching using AssistEM. *Int. J. Data Sci. Anal.* **2026**, *22*, 255. https://doi.org/10.1007/s41060-026-01223-5.
175. Napoleone, M.; Console, M.; Lenzerini, M.; Merialdo, P.; Papi, L.; Poggi, A.; Scafoglieri, F.; Torlone, R. Domain-Constrained Data Augmentation for Entity Resolution. In *Communications in Computer and Information Science*; Springer: Cham, Switzerland, 2026; pp. 554–566. https://doi.org/10.1007/978-3-032-19096-3_37.
176. Zhao, W.X.; Zhou, K.; Li, J.; Tang, T.; Dong, Z.; Hou, Y.; Zhang, B.; Min, Y.; Zhang, J.; Liu, P.; et al A Survey of Large Language Models. *Front. Comput. Sci.* **2026**, *20*, 2012627. https://doi.org/10.1007/s11704-026-60308-3.
177. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training Language Models to Follow Instructions with Human Feedback. In *Proceedings of the Advances in Neural Information Processing Systems*; ICLR: Appleton, WI, USA, 2022; Volume 35.
178. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Ichter, B.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In *Proceedings of the Advances in Neural Information Processing Systems*; ICLR: Appleton, WI, USA, 2022; Volume 35.
179. Hu, E.J.; Shen, Y.; Wallis, P.; Allen-Zhu, Z.; Li, Y.; Wang, S.; Wang, L.; Chen, W. LoRA: Low-Rank Adaptation of Large Language Models. In *Proceedings of the International Conference on Learning Representations (ICLR)*; NeurIPS: San Diego, CA, USA, 2022.
180. Beheshti, M.; Gondara, L.; Zachary, I. Leveraging Language Models for Automated Patient Record Linkage. *arXiv* **2025**, arXiv:2504.15261. https://doi.org/10.48550/arXiv.2504.15261.