

Jul 17, 2025

Bibliometric Review Protocol: The Road to Autonomy – AI in Autonomous Vehicles

DOI

<https://dx.doi.org/10.17504/protocols.io.e6nvw49m2lmk/v1>

Camelia Delcea¹

¹Bucharest University of Economic Studies



Camelia Delcea

Bucharest University of Economic Studies

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.e6nvw49m2lmk/v1>

Protocol Citation: Camelia Delcea 2025. Bibliometric Review Protocol: The Road to Autonomy – AI in Autonomous Vehicles. protocols.io <https://dx.doi.org/10.17504/protocols.io.e6nvw49m2lmk/v1>

License: This is an open access protocol distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

We use this protocol and it's working

Created: July 17, 2025

Last Modified: July 17, 2025

Protocol Integer ID: 222662

Keywords: automono*_electric_vehicle*, machine_learning, deep_learning, artificial_intelligence, bibliometric review protocol, bibliometric review, ai in autonomous vehicle, ai, autonomous vehicle, quality publication, road to autonomy, quality publications from the web, artificial intelligence, intersection of artificial intelligence, deep learning, autonomy, machine learning

Abstract

This protocol outlines the methodological framework for a bibliometric review focused on the intersection of artificial intelligence (AI), machine learning (ML), deep learning (DL), and autonomous vehicles (AV). The aim is to map academic production, collaboration patterns, thematic trends, and emerging technologies by analyzing high-quality publications from the Web of Science (WoS) Core Collection.

Guidelines

Eligibility Criteria

Inclusion:

- Language: English only
- Document type: Article
- Publication year: 1995–2024
- Content: Must include all four domains (AV, AI, ML, DL)

Exclusion:

- Non-English papers (29)
- Non-article types (1769)
- Articles published in 2025 (63)
- Manually excluded irrelevant topics (37)

Final dataset: 2,228 articles.

Screening Process

- PRISMA 2020 methodology was followed.
- Articles were screened manually after keyword filtering.
- Duplicate and off-topic articles were removed.
- Figure 1 in the paper presents the PRISMA flow diagram.

Tools and Software

- **Biblioshiny 5.0 in R Studio 4.3.2** (for bibliometric analyses)
- **Gensim (Python) for Latent Dirichlet Allocation (LDA)**
- **BERTopic (Python)** using **HDBSCAN** clustering

Data Preprocessing for NLP

- Text was lowercased
- Punctuation removed
- Token normalization applied (e.g., “AI” → “artificial intelligence”)
- Domain-specific stopwords removed
- Grid search applied for LDA alpha/eta tuning
- BERTopic clustering tuned via minimum cluster/sample sizes

Key Outcomes Measured

- Publication volume by year and country
- Most cited articles and normalized total citations (NTC)
- Bradford's Law core journals and H-index analysis
- Thematic evolution and keyword co-occurrence networks
- Topic modeling with LDA and BERTopic



- Collaboration maps for authors and countries

Materials

Web of Science (WoS) Core Collection institutional subscription; Biblioshiny 5.0 in R Studio 4.3.2; Gensim (Python); BERTopic (Python); Computer with internet access.

Before start

Special search symbols:

- _ was used to capture exact multi-word phrases.
- * wildcard operator was used for plural/singular and extended term variations.

Databases and Indexes Used

- 1 Extract all articles from the **Web of Science (WoS) Core Collection** using an institutional subscription.
Include the following indexes in the extraction:
 - Science Citation Index Expanded (SCIE)
 - Social Sciences Citation Index (SSCI)
 - Arts and Humanities Citation Index (A&HCI)
 - Emerging Sources Citation Index (ESCI)
 - Conference Proceedings Citation Index – Science (CPCI-S)
 - CPCI – Social Sciences and Humanities (CPCI-SSH)
 - Book Citation Index – Science (BKCI-S)
 - Book Citation Index – SSH (BKCI-SSH)
 - Index Chemicus (IC)
 - Current Chemical Reactions (CCR-Expanded)

Search Strategy

- 2 Apply the following keyword queries in the Title (TI), Abstract (AB), and Author Keywords (AK) fields:
 - "autonomo*electricvehicle"
 - "machine_learning"
 - "deep_learning"
 - "artificial_intelligence"
- 3 Use special search symbols:
 - _ to capture exact multi-word phrases.
 - * as a wildcard operator for plural/singular and extended term variations.

Eligibility Criteria

- 4 Apply the following inclusion criteria:
 - Language: English only
 - Document type: Article
 - Publication year: 1995–2024
 - Content: Must include all four domains (AV, AI, ML, DL)
- 5 Apply the following exclusion criteria:
 - Non-English papers (29)
 - Non-article types (1769)
 - Articles published in 2025 (63)
 - Manually excluded irrelevant topics (37)



- 6 Finalize the dataset to include 2,228 articles after applying inclusion and exclusion criteria.

Screening Process

- 7 Follow PRISMA 2020 methodology for screening.
- 8 Screen articles manually after keyword filtering.
- 9 Remove duplicate and off-topic articles from the dataset.
- 10 Refer to Figure 1 in the paper for the PRISMA flow diagram.

Tools and Software

- 11 Use **Biblioshiny 5.0** in **R Studio 4.3.2** for bibliometric analyses.
- 12 Use **Gensim (Python)** for **Latent Dirichlet Allocation (LDA)** topic modeling.
- 13 Use **BERTopic (Python)** with **HDBSCAN** clustering for topic modeling.

Data Preprocessing for NLP

- 14 Lowercase all text data.
- 15 Remove punctuation from text data.
- 16 Apply token normalization (e.g., convert "AI" to "artificial intelligence").



- 17 Remove domain-specific stopwords from the text.
- 18 Apply grid search for LDA alpha/eta parameter tuning.
- 19 Tune BERTopic clustering using minimum cluster and sample sizes.

Key Outcomes Measured

- 20 Measure publication volume by year and country.
- 21 Identify most cited articles and calculate normalized total citations (NTC).
- 22 Analyze Bradford's Law core journals and perform H-index analysis.
- 23 Map thematic evolution and keyword co-occurrence networks.
- 24 Conduct topic modeling using LDA and BERTopic.
- 25 Generate collaboration maps for authors and countries.

Acknowledgements

Retrospective Registration Note: This protocol was registered **after the start of data extraction**, in line with transparency practices recommended by PRISMA 2020.