Islamic Knowledge Management
EISSN: 2476-7255

# Semantic Vector Representation: An Introduction to Its Foundations and Applications in the Humanities*

**Ammar Jokar**[1] **Mohammad Hadi Shariati Firouzabad**[2]

1. PhD Candidate in Culture and Communication, Baqir al-Olum University, Qom, Iran (**Corresponding Author**). ajokar@iran.ir

2. Assistant Professor, Department of Cultural Policy, Baqir al-Olum University, Qom, Iran. m.hadi.shariati@gmail.com

## Abstract

This study aims to elucidate the theoretical foundations of semantic vector representation and introduce its potential for textual analysis, the discovery of semantic relationships, and the development of research methods in the humanities. The central research problem lies in the inability of machines to directly comprehend natural language, coupled with the relative limitations of traditional and classical digital methods in analyzing large-scale textual data. Adopting a descriptive–analytical approach and drawing on a review of established works in natural language processing and computational linguistics, the study examines the foundations of semantic vector representation, the distributional hypothesis of meaning, and its role in the computational modeling of language. The findings indicate that by transforming words and texts into multidimensional spaces, vector representations

* Jokar, A., & Shariati Firouzabad, M.-H. (2025). Semantic vector representation: An introduction to its foundations and applications in the humanities. *Islamic Knowledge Management, 7*(2), pp. 166-195. https://Doi.org/10.22081/jikm.2026.75210.1128

▣ **Article Type:** Research; **Publisher:** Islamic Sciences and Culture Academy, Qom, Iran

▣ **Received:** 2025/05/26 ● **Revised:** 2025/08/18 ● **Accepted:** 2025/09/08 ● **Published online:** 2025/10/01

enable the measurement of semantic similarity, analysis of conceptual relationships, tracking of the historical evolution of concepts, extraction of discourse clusters, and development of semantic search in humanities texts. Nevertheless, because such models lack lived experience, intentionality, and consciousness, they cannot replace human understanding and interpretation; rather, they function as complementary tools for expanding research and knowledge management capabilities in the humanities.

**Keywords**
Semantic Vector Representation, Large Language Models, Distributional Hypothesis of Meaning, Word Embeddings, Natural Language Processing, Semantic Space, Language Modeling.

إدارة المعرفة الإسلامية
EISSN: ٢٤٧٦-٧٢٥٥

# التمثيل المتجهي للمعنى:
# مدخل إلى أساسياته وتطبيقاته في العلوم الإنسانية*

عمار جوكار[١] محمدهادي شريعتي فيروزآباد[٢]
١.طالب دكتوراه في الثقافة والعلاقات، جامعة باقر العلوم عليه السلام، قم، إيران (**الباحث المسؤول**).
ajokar@iran.ir
٢.أستاذ مساعد في قسم السياسات الثقافية، جامعة باقر العلوم عليه السلام، قم، إيران.
m.hadi.shariati@gmail.com



**الملخص**

تهدف هذه الدراسة إلى شرح الأسس النظرية للتمثيل المتجهي للمعنى، واستعراض إمكانياته في تحليل النصوص، واكتشاف العلاقات الدلالية، وتطوير مناهج البحث في العلوم الإنسانية. تنبثق المسألة المركزية للدراسة من عجز الآلة عن الفهم المباشر للغة الطبيعية، وفي المقابل، القصور النسبي للطرائق الرقمية التقليدية والكلاسيكية في التعامل مع البيانات النصية الضخمة وتحليلها. تتناول هذه الدراسة، من خلال منهج وصفي تحليلي ومراجعة للأعمال المرجعية في مجال معالجة اللغة الطبيعية واللغويات الحاسوبية مبادئ التمثيل المتجهي للمعنى، و«الفرضية التوزيعية للمعنى» ودورها في النمذجة

* **الاستشهاد بهذه المقالة:** جوكار، عمار، محمدهادي، شريعتي فيروزآباد. (١٤٠٤). التمثيل المتجهي للمعنى: مدخل إلى أساسياته وتطبيقاته في العلوم الإنسانية. إدارة المعرفة الإسلامية، ٧(٢)، ص ١٦٦-١٩٥.
https://Doi.org/10.22081/jikm.2026.75210.1128

▣ **نوع المقال:** بحثية، **الناشر:** المعهد العالي للعلوم والثقافة الإسلامية، قم، إيران

▣ **تاريخ الإستلام:** ٢٠٢٥/٠٤/٢٨ ● **تاريخ التعديل:** ٢٠٢٥/٠٧/١٨ ● **تاريخ القبول:** ٢٠٢٥/٠٨/٠١ ● **تاريخ الإصدار:** ٢٠٢٥/١٠/٠١

الحاسوبية للغة. تُظهر النتائج أن التمثيل المتجهي عبر تحويل الكلمات والنصوص إلى فضاءات رياضية متعددة الأبعاد يتيح قياس التشابه الدلالي، وتحليل الروابط المفاهيمية، وتتبع التحولات التاريخية للمفاهيم، واستخراج العناقيد الخطابية، وتطوير البحث الدلالي في نصوص العلوم الإنسانية. ومع ذلك، فإن هذه النماذج الحاسوبية، لافتقارها إلى الخبرة المعاشة والقصدية والوعي، لا تشكل بديلاً عن الفهم والتأويل الإنساني، بل تعمل بوصفها أداة مكملة تهدف إلى توسيع آفاق البحث وإثراء نظم إدارة المعرفة في حقل العلوم الإنسانية.

**الكلمات المفتاحية**

التمثيل المتجهي للمعنى، النماذج اللغوية الكبيرة، فرضية توزيع المعنى، تضمين الكلمات، معالجة اللغة الطبيعية، الفضاء الدلالي، نمذجة اللغة.

مدیریت دانش اسلامی
EISSN: ۲۴۷۶-۷۲۵۵

# بازنمایی برداری معنا:
# درآمدی بر مبانی و کاربردها در علوم انسانی*

عمار جوکار[۱] محمدهادی شریعتی فیروزآباد[۲]

۱. دانشجوی دکتری فرهنگ و ارتباطات دانشگاه باقرالعلوم(ع)، قم. ایران (**نویسنده مسئول**).
ajokar@iran.ir

۲. استادیار گروه سیاست‌گذاری فرهنگی دانشگاه باقرالعلوم(ع)، قم. ایران.
m.hadi.shariati@gmail.com



## چکیده

پژوهش حاضر با هدف تبیین مبانی نظری بازنمایی برداری معنا و معرفی ظرفیت‌های آن در تحلیل متون، کشف روابط معنایی و توسعه روش‌های پژوهش در علوم انسانی است. مسئله اصلی پژوهش، ناتوانی ماشین در فهم مستقیم زبان طبیعی و در مقابل، ناکارآمدی نسبی روش‌های سنتی و کلاسیک دیجیتال در تحلیل کلان‌داده‌های متنی است. این مطالعه با رویکرد توصیفی - تحلیلی و بر پایه مرور آثار معتبر در حوزه پردازش زبان طبیعی و زبان‌شناسی محاسباتی، مبانی بازنمایی برداری معنا، فرضیه توزیعی معنا و نقش آن در مدل‌سازی محاسباتی زبان را بررسی می‌کند. یافته‌ها نشان می‌دهد که بازنمایی برداری با تبدیل واژه‌ها و متون به فضاهای چندبعدی، امکان سنجش شباهت معنایی، تحلیل روابط مفهومی، ردیابی تحول تاریخی مفاهیم، استخراج خوشه‌های گفتمانی و توسعه جست‌وجوی معنایی را در متون علوم انسانی فراهم می‌سازد. با وجود این، چنین مدل‌هایی

* **استناد به این مقاله:** جوکار، عمار؛ شریعتی فیروزآباد، محمدهادی. (۱۴۰۴). بازنمایی برداری معنا: درآمدی بر مبانی و کاربردها در علوم انسانی. مدیریت دانش اسلامی، ۷(۲)، صص ۱۶۶-۱۹۵.
https://Doi.org/10.22081/jikm.2026.75210.1128

▣ **نوع مقاله:** پژوهشی؛ **ناشر:** پژوهشگاه علوم و فرهنگ اسلامی، قم، ایران

▣ **تاریخ دریافت:** ۱۴۰۴/۰۳/۰۵ ● **تاریخ اصلاح:** ۱۴۰۴/۰۵/۲۷ ● **تاریخ پذیرش:** ۱۴۰۴/۰۶/۱۷ ● **تاریخ انتشار آنلاین:** ۱۴۰۴/۰۷/۰۹

به دلیل فقدان تجربه زیسته، قصدمندی و آگاهی، جایگزین فهم و تفسیر انسانی نیستند، بلکه به‌عنوان ابزاری مکمل برای گسترش ظرفیت‌های پژوهش و مدیریت دانش در علوم انسانی عمل می‌کنند.


## کلیدواژه‌ها

بازنمایی برداری معنا، مدل‌های زبانی بزرگ، فرضیه توزیعی معنا، تعبیه‌سازی کلمات، پردازش زبان طبیعی، فضای معنایی، مدل‌سازی زبان.

## مقدمه و بیان مسئله

در دهه‌های اخیر، گسترش فناوری‌های دیجیتال موجب تولید و انباشت حجم عظیمی از داده‌های متنی در حوزه‌های علوم انسانی و اسلامی شده است. کتاب‌ها، مقالات، اسناد تاریخی، منابع تفسیری، حدیثی، فقهی و داده‌های منتشرشده در فضای مجازی، مجموعه‌ای گسترده از کلان‌داده‌های متنی را شکل داده‌اند که ظرفیت ارزشمندی برای تولید دانش و کشف الگوهای جدید دارند. با این حال، روش‌های سنتی پژوهش و حتی شیوه‌های کلاسیک دیجیتال که عمدتاً بر جست‌وجوی واژگانی، نمایه‌سازی دستی و تحلیل‌های محدود متکی هستند، توانایی کافی برای تحلیل و تقاطع‌گیری میان این حجم گسترده از داده‌ها را ندارند. درنتیجه، بسیاری از روابط پنهان، الگوهای کلان و بینش‌های نوظهور در میان انبوه متون ناشناخته باقی می‌مانند.

در چنین شرایطی، ظهور هوش مصنوعی و بویژه پیشرفت‌های چشمگیر در حوزه پردازش زبان طبیعی[1]، افق‌های تازه‌ای را پیش‌روی پژوهشگران گشوده است. این فناوری‌ها امکان تحلیل خودکار متون، استخراج دانش و برقراری پیوند میان منابع پراکنده را فراهم کرده‌اند. با وجود این، یکی از مهمترین چالش‌های پیش‌روی این فناوری‌ها، مسئله درک معنا توسط ماشین است. از آنجاکه علوم انسانی و اسلامی اساساً با مفاهیم، معانی، تفاسیر و روابط معنایی سروکار دارند، صرف شناسایی واژگان یا الگوهای آماری برای دستیابی به فهم عمیق متون کافی نیست و لازم است ماشین بتواند ساختارهای معنایی نهفته در زبان را درک کند.

در پاسخ به این چالش، رویکردهای گوناگونی برای بازنمایی زبان[2] توسعه یافته است که هر یک می‌کوشند زبان طبیعی را به‌صورتی قابل پردازش برای ماشین تبدیل کنند. در میان این رویکردها، بازنمایی برداری معنا[3] به دلیل توانایی در مدل‌سازی روابط معنایی و فراهم کردن بستری برای تحلیل مفاهیم، نقطه عطفی در تحول پردازش زبان

1. Natural Language Processing (NLP)
2. Language Representation
3. Distributed Semantic Representation

طبیعی به شمار می‌رود. از این‌رو، مقاله حاضر می‌کوشد ضمن تبیین مبانی بازنمایی برداری معنا، نشان دهد که این رویکرد چگونه می‌تواند پلی میان فناوری‌های نوین پردازش زبان و مسائل پژوهشی علوم انسانی و اسلامی برقرار کند و زمینه تحلیل معنایی کلان‌داده‌های متنی را فراهم آورد.

## پیشینه پژوهش

پژوهش‌های مرتبط با بازنمایی برداری معنا را می‌توان در سه محور اصلی دسته‌بندی کرد: محور نخست، پژوهش‌های فنی درباره مدل‌های برداری واژگان است که بر توانایی مدل‌های برداری در استخراج روابط معنایی از هم‌نشینی واژگان تأکید دارند. در این زمینه، وِورس[1] و کولن[2] نشان دادند که مدل‌های برداری کلمات[3] قادرند تغییرات معنایی[4] را در متون تاریخی آشکار سازند و امکان تحلیل پویای معنا را فراهم کنند (وورس و کولن، ۲۰۲۰، صص ۲۲۶–۲۲۷). محور دوم، مطالعات علوم انسانی در پیکره‌های محدود است. پژوهش‌های وُلگِنَنت[5] و همکاران نشان می‌دهد که حتی در پیکره‌های محدود علوم انسانی نیز مدل‌های برداری‌ای مانند وردتووِک[6]، گلاو[7] و فست‌تکست[8] توان استخراج روابط معنایی و قیاسی را دارند و می‌توان از آنها برای تحلیل متون ادبی و فرهنگی بهره گرفت (ولگنت و دیگران، ۲۰۱۹، صص ۱–۲). همچنین، همیلتون و همکارانش با استفاده از بردارهای تاریخی نشان دادند که تغییر معنای واژگان تابع الگوهای آماری قابل اندازه‌گیری است (همیلتون و دیگران، ۲۰۱۶، ص ۹). محور سوم، کاربردهای معناشناسی برداری[9] در متون دینی و اسلامی است. اگرچه پژوهش‌ها در این حوزه هنوز در ابتدای راه قرار

1. Wevers
2. Koolen
3. Word Embeddings Models
4. Semantic Shifts
5. Wohlgenannt
6. Word2Vec
7. GloVe
8. FastText
9. Vectorized Semantics

دارد، اما نمونه‌هایی مانند کاربست بازنمایی برداری متن قرآن، جهت بهبود جستجوی محتوای آن(محمد و شکری، ۲۰۲۲، ص ۱) بیان‌گر ظرفیت مدل‌های برداری برای تحلیل مفاهیم قرآنی، شبکه‌های معنایی حدیث و تحلیل گفتمان دینی است. ازاین‌رو، پژوهش حاضر می‌کوشد پس از تبیین مفاهیم مرتبط با بازنمایی برداری معنا، ظرفیت‌های کاربردی آن در مطالعات علوم انسانی و اسلامی را توضیح دهد.

## مفاهیم و مبانی

### ۱. چالش درک معنا برای ماشین

چالش ماشین در فهم زبان و معنا یکی از بنیادی‌ترین مسائل در تقاطع علوم رایانه، زبان‌شناسی و علوم شناختی است. در نگاه نخست، عملکرد سامانه‌های هوش مصنوعی در ترجمه ماشینی، پاسخ‌گویی به پرسش‌ها و تولید متن، این تصور را ایجاد می‌کند که ماشین نیز همانند انسان زبان را درک می‌کند؛ اما این شباهت رفتاری لزوماً به معنای فهم واقعی معنا نیست. زبان انسانی تنها مجموعه‌ای از واژه‌ها و قواعد دستوری نیست، بلکه نظامی پیچیده است که مفاهیم، تجربه‌های زیسته، زمینه‌های فرهنگی، قصد گوینده و روابط اجتماعی را در خود جای داده است. هنگامی که انسان جمله‌ای را درک می‌کند، تنها به واژه‌های موجود در آن توجه ندارد، بلکه با بهره‌گیری از دانش پیشین، تجربه، زمینه موقعیتی و مدل ذهنی خود از جهان، معنای آن را تفسیر می‌کند. ازاین‌رو، فهم زبان برای انسان فرآیندی شناختی و مفهومی است که فراتر از تحلیل صرف واژه‌ها و ساختارهای زبانی قرار دارد.

در مقابل، ماشین فاقد تجربه زیسته، دانش شهودی و ارتباط مستقیم با جهان خارج است. رایانه جهان را نه به‌صورت مفهومی، بلکه به‌صورت داده‌های قابل محاسبه پردازش می‌کند و متن برای آن چیزی جز دنباله‌ای از نمادها نیست که باید به قالبی مناسب برای انجام محاسبات تبدیل شوند. به همین دلیل، حتی پیشرفته‌ترین مدل‌های زبانی نیز، با وجود تولید متونی روان و منسجم، در مواردی مانند درک کنایه، استعاره، ارجاعات فرهنگی یا استدلال‌های وابسته به زمینه با محدودیت‌هایی روبه‌رو هستند.

مریل[1] و همکاران تأکید می‌کنند که مدل‌هایی که تنها براساس فرم متنی آموزش می‌بینند، به دلیل نداشتن پیوند مستقیم با جهان واقعی و تجربه انسانی، با محدودیت‌های بنیادین در درک معنا مواجه‌اند(مریل و دیگران، ۲۰۲۱، ص ۱۰۴۷). همچنین، میچل و کراکاور نشان می‌دهند که مدل‌های زبانی بزرگ[2] پیش‌آموزش‌دیده مانند جی. پی. تی۳[3] و لامدا[4]، صرف‌نظر از روانی زبان تولیدی، فاقد درک واقعی هستند؛ زیرا این مدل‌ها تجربه یا مدل ذهنی از جهان ندارند و تنها از طریق پیش‌بینی کلمات در مجموعه‌های بزرگ داده‌های متنی، ساختار زبان را یاد می‌گیرند، اما معنای آن را درک نمی‌کنند(میچل و کراکاور، ۲۰۲۳، صص ۲–۳).

از منظر علوم رایانه نیز این محدودیت ریشه در ماهیت محاسباتی ماشین دارد. زبان طبیعی یک نظام نمادین گسسته است، در حالی که رایانه تنها قادر به انجام عملیات بر روی داده‌های محاسباتی است. فرون[5] و زانزوتتو[6] توضیح می‌دهند که اگرچه زبان انسانی ذاتاً یک نظام نمادین است، مدل‌های یادگیری ماشین برای انجام محاسبات، آن را به بازنمایی‌های پیوسته تبدیل می‌کنند(فرون و زانزوتتو، ۲۰۲۰: ص ۲). بنابراین، پرسش اساسی این است که چگونه می‌توان زبان و روابط معنایی نهفته در آن را به گونه‌ای بازنمایی کرد که هم برای ماشین قابل پردازش باشد و هم تا حد امکان اطلاعات معنایی متن را حفظ کند؟ پاسخ به این پرسش، زمینه‌ساز شکل‌گیری رویکردهای نوینی در بازنمایی زبان، بویژه بازنمایی برداری معنا شده است که در ادامه به آن پرداخته خواهد شد.

### ۲. ایده بازنمایی برداری معنا

یکی از مهمترین پاسخ‌های هوش مصنوعی به چالش درک زبان و معنا، بازنمایی

1. Merrill
2. Large Language Models(LLMs)
3. GPT-3
4. LaMDA
5. Ferrone
6. Zanzotto

بـرداری اسـت. از جملـه تحـولات اساسـی در پـردازش زبـان طبیعـی، تبـدیل واژه‌هـا از واحدهای زبانی به واحدهای محاسباتی است؛ فرآیندی که به‌عنوان تعبیه‌سازی کلمـات[1] به کمک بردارهای عددی شناخته می‌شود. در این رویکـرد، هـر واژه بجـای آنکـه تنهـا یک نشانه زبانی باشد، به یک بردار در یک فضای چندبعدی تبدیل می‌شود که موقعیت نسبی آن را در میـان سـایر واژه‌هـا تعیـین می‌کنـد. میکولـوف[2] و همکـاران در پـژوهش مشهور خود نشان دادند که مدل‌های برداری قادرند روابط معنایی و نحوی میـان واژه‌هـا را از طریق الگوهای هم‌رخدادی در متن کشف کنند(میکولـوف و دیگـران، ۲۰۱۳: ص ۲). ایـن تحول، زمینه را برای توسعه نسل جدیدی از روش‌های پردازش زبان طبیعی فراهم کرد.

ایده مرکزی این رویکرد بر یک فرض ساده اما بسیار قدرتمند استوار اسـت: معنـا از زمینه استعمال واژه‌ها پدید می‌آیـد. ایـن اصـل کـه بـه فرضـیه تـوزیعی[3] معـروف اسـت، نخستین بار در زبان‌شناسی توصیفی مطرح شد و بعدها مبنای اصلی بسـیاری از مـدل‌های محاسباتی زبان قرار گرفـت. براسـاس ایـن دیـدگاه، معنـا نـه در خـود واژه‌هـا، بلکـه در الگوهای هم‌نشینی آنها در متن شـکل می‌گیـرد. (جورافسـکی و مـارتین، ۲۰۲۶، ص ۹۶) بـه بیـان ساده، اگر دو واژه در جملات مشابه یا با هم‌واژه‌های مشابه ظاهر شـوند، احتمـالاً از نظـر معنایی به هم نزدیک‌اند. نکته مهم در اینجا آن است که برداری‌سازی، معنا را جـایگزین نمی‌کند، بلکه آن را مدل‌سازی[4] می‌کند. این نکتـه بـرای علـوم انسـانی اهمیـت ویـژه‌ای دارد، زیرا مانع از ساده‌سازی بیش از حد مفهوم معنا در تحلیل‌هـای محاسـباتی می‌شـود. این مسئله، نوعی چارچوب به ما می‌دهد که در آن، روابط میان کلمـات بـه انـدازه خـود کلمات در یک بافت اهمیت دارند.

در مدل‌های بـرداری، ایـن ایـده بـه شـکلی محاسـباتی پیاده‌سـازی می‌شـود. پـس از آموزش مدل بر روی مجموعه‌های بزرگ متنی[5]، این بردارهـا قـادر بـه بازنمـایی روابـط

1. Word Embeddings
2. Mikolov
3. Distributional Hypothesis
4. Modeling
5. Datasets

معنایی پیچیده بین کلمات هستند، از جمله روابط جغرافیایی مانند نسبت بین یک کشور و شهر متعلق به آن، به‌طوری‌که نسبت بردار «فرانسه» به «پاریس» مشابه نسبت بردار «آلمان» به «برلین» است(میکولوف و دیگران، ۲۰۱۳، ص ۵). همچنین آلمیدا[1] و ششیو[2] توضیح می‌دهند که بردارها می‌توانند اطلاعات معنایی، نحوی و ارتباطی واژه‌ها را در بردارهایی متراکم ذخیره کنند(المیدا و ششیو، ص ۱). این توانایی، امکان معناشناسی برداری[3] را به ما می‌دهد و می‌توان به‌کمک آن، فضای معنایی[4] چندبعدی را ترسیم کرد (جورافسکی و مارتین، ۲۰۲۶، ص ۱۰۰). با این حال، بازنمایی برداری از همان ابتدا به شکل امروزی وجود نداشت، بلکه حاصل چند دهه تحول در شیوه‌های بازنمایی زبان در پردازش زبان طبیعی است. بررسی این روند نشان می‌دهد که چگونه پژوهشگران گام‌به‌گام از بازنمایی‌های ساده واژگانی به مدل‌های پیچیده و زمینه‌محور امروزی رسیده‌اند.



### ۳. روش‌های بازنمایی زبان

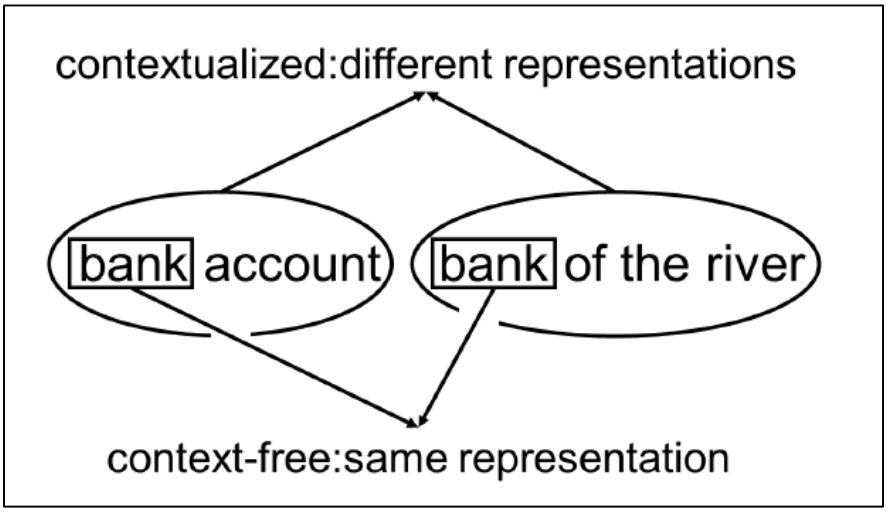
سیر تحول بازنمایی زبان در پردازش زبان طبیعی را می‌توان به سه گام اساسی و چند معماری کلیدی تقسیم کرد که نشان می‌دهند چگونه زبان از واحدهای نمادین به بازنمایی‌های برداری و زمینه‌محور تبدیل شده است:

#### گام اول: بازنمایی کلمات و مسئله معنای بافتارمحور

نخستین رویکرد جدی در بازنمایی محاسباتی زبان، ابداع مدل‌های تعبیه‌سازی کلمات بود. مدل‌های اولیه – مانند وُردتووِک – کلمات را به‌صورت ایستا یا فارغ از بافت[5] متن کدگذاری می‌کردند؛ به این معنا که یک واژه چندمعنایی در تمام جملات،

1. Almeida
2. Xexéo
3. Vector semantics
4. Semantic space
5. context-free

ارزش یکسانی داشت. از منظر معناشناختی، این روش اثربخش نبود، چراکه معنا در شبکه روابط واژگان شکل می‌گیرد. نسل بعدی این مدل‌ها با معرفی بازنمایی‌های بافت‌محور[1] این شکاف را پر کردند. در این رویکرد، ارزش محاسباتی یک واژه براساس کلمات هم‌جوار آن تعیین می‌شود؛ ساختاری که معنای زبان را نه به‌صورت مجرد، بلکه در ظرف محیطی و بافت متن تحلیل می‌کند (شوماخر و تروپمن-فریک، ۲۰۲۱، ص ۴). همان‌گونه که در تصویر (۱) مشاهده می‌شود، واژه bank – که در زبان انگلیسی هم در معنای ساحل و هم در معنای بانک به‌کار برده می‌شود – در رویکرد فارغ از بافت، در دو جمله متفاوت، بازنمایی یکسانی دارد و به عبارت دیگر، ماشین تفاوت معنای این کلمه در بافت را درک نمی‌کند. اما در بازنمایی بافت‌محور، ماشین برای همین لفظ، در دو جمله مختلف با توجه به کلمات مجاور آن، بازنمایی متفاوتی را ایجاد می‌کند و از حیث معنایی، میان آنها تمایز قایل می‌شود.




**تصویر (۱) – تفاوت بازنمایی کلمه bank در دو رویکرد بافت‌محور و فارغ از بافت**

1. Contextualized Representations

### گام دوم: الگوریتم‌های توالی‌محور و فرآیند انتقال معنا

زبان، ماهیتی زمان‌مند و توالی‌محور دارد و معنای یک گزاره با تقدم و تاخر کلمات تغییر می‌کند. برای شبیه‌سازی این ویژگی گرامری، شبکه‌های عصبی بازگشتی [1]طراحی شدند. این مدل‌ها با ایجاد یک جریان حافظه، اطلاعات کلمات قبلی را به کلمات بعدی منتقل می‌کنند تا ساختار خطی زبان حفظ شود. با این حال، در حوزه‌هایی مانند ترجمه یا تلخیص متن، نیاز به بازسازی کل ساختار جمله بود که منجر به ابداع معماری رمزگذار-رمزگشا[2] شد. این ساختار فرآیند فهم و تولید زبان را به دو بخش تقسیم می‌کند: بخش رمزگذار ابتدا کل متن ورودی را تحلیل و عصاره معنایی آن را در یک بردار مفهومی فشرده می‌کند و سپس بخش رمزگشا براساس این هویت انتزاعی، متن جدید را تولید می‌نماید (شوماخر و تروپمان-فریک، ۲۰۲۱، ص ۴ و ۵). این فرآیند شباهت ساختاری بالایی با مدل‌های کلاسیک فرآیند ارتباطات، رمزگذاری مفاهیم و انتقال پیام دارد.

### گام سوم: مکانیسم توجه و انقلاب پردازش موازی

محدودیت بزرگ مدل‌های توالی‌محور، ضعف حافظه در رویارویی با متون طولانی بود، زیرا فشرده‌سازی یک متن بلند در یک بردار ثابت، موجب از بین رفتن جزئیات نظری و معنایی می‌شد. ابداع مکانیسم توجه[3] این چالش را برطرف ساخت. این الگوریتم به ماشین امکان می‌دهد که هنگام پردازش یک بخش از متن، به‌طور هم‌زمان به تمام اجزای دیگر ورودی دسترسی داشته باشد و روی بخش‌های کلیدی تمرکز انتخابی کند. با تکیه بر این مکانیسم، معماری پایه ترانسفورمر[4] متولد شد. ترانسفورمرها پردازش خطی و کلمه به کلمه را کاملاً کنار گذاشتند و به پردازش

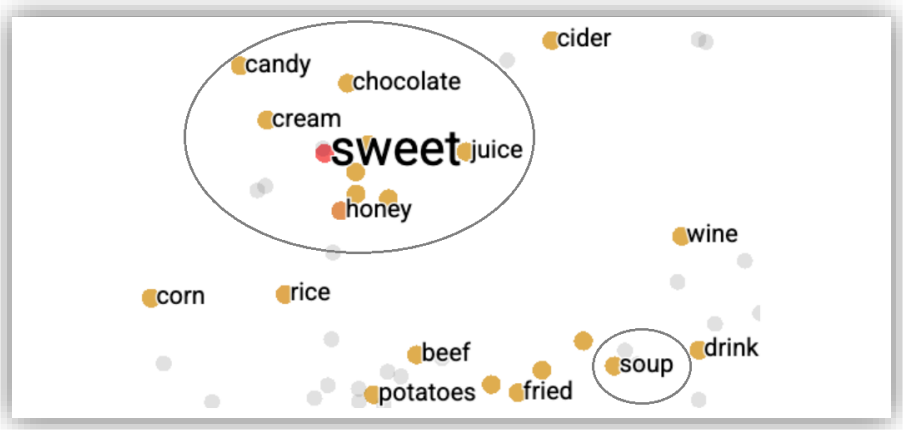
1. Recurrent Neural Networks (RNNs)
2. Encoder-Decoder
3. Attention
4. Transformer

نظام‌مند و موازی کل متن رو آوردند(شوماخ و تروپمان-فریک، ۲۰۲۱، صص ۵ و ۶). این ویژگی به ماشین اجازه می‌دهد ارتباطات پنهان و ساختاری میان تمام عناصر یک متن را کشف کند.

این سیر تحول در رویکردهای بازنمایی زبان نشان می‌دهد که بردارهای معنایی را می‌توان به‌عنوان زبان مشترک بسیاری از مدل‌های زبانی در نظر گرفت. در این چارچوب، لایه‌های مختلف تحلیل زبانی، از مدل‌های ابتدایی تا ساختارهای پیچیده‌تر، همگی در پی نمایش روابط میان مفاهیم در قالب ساختارهای قابل محاسبه هستند. چنین رویکردی امکان عبور از سطح واژه‌ها و توجه به پیوندهای معانی را فراهم می‌کند. با این حال، کارکرد اصلی این بازنمایی‌ها فقط نمایش نیست، بلکه ایجاد امکان سنجش فاصله و شباهت در یک فضای معنایی مشترک است. درنتیجه، بازنمایی برداری را می‌توان نقطه اتصال رویکردهای متنوع دانست که امکانِ گذار از شباهت‌های سطحی به سازمان‌دهی ساختاری معانی را فراهم می‌کند. این پرسش در اینجا برجسته می‌شود که این فضای برداری چگونه به تشخیص شباهت معنایی و در ادامه به‌صورت‌بندی نقشه‌های معنایی منجر می‌شود؟ پاسخ به این پرسش، موضوع بخش بعدی است.

### ۴. از کشف شباهت معنایی تا نقشه معنایی

برای انسان، نزدیکی معنایی میان واژه‌هایی مانند «استاد» و «دانشجو» در مقایسه با «درخت» بدیهی است، اما برای ماشین چنین ادراکی از پیش وجود ندارد. همان‌گونه که پیش‌تر اشاره شد، مدل‌های برداری بر فرضیه توزیعی معنا استوارند و شباهت واژه‌ها را از طریق نزدیکی آنها در فضای برداری محاسبه می‌کنند. در این چارچوب، هر واژه به‌صورت یک بردار عددی در نظر گرفته می‌شود که موقعیت آن حاصل الگوهای هم‌رخدادی در متن است؛ بنابراین، واژه‌هایی که در بافت‌های مشابه ظاهر می‌شوند، در این فضا به یکدیگ نزدیک‌تر قرار می‌گیرند.




**تصویر(۲): بازنمایی برداری فضای معنایی واژه «شیرین(Sweet)»**

به‌عنوان مثال، با توجه به تصویر(۲) در فضای معنایی واژه «شیرین[1]»، واژه‌هایی مانند «عسل[2]»، «خامه[3]» و «شکلات[4]» معمولاً در کنار همین کلمه در متون مربوط به خوراکی‌ها ظاهر می‌شوند، در حالی که واژه‌ای مانند «سوپ[5]» یا «سیب‌زمینی[6]» در چنین بافت‌هایی در فضای معنایی واژه «شیرین» قرار نمی‌گیرد و از آن فاصله بیشتری دارند. نتیجه این فرایند آن است که بردارهای کلمه «شیرین» و «عسل» در فضای معنایی به هم نزدیک‌تر از «شیرین» و «سوپ» قرار می‌گیرند (جورافسکی و مارتین، ۲۰۲۶: ص ۱۰۰). این الگوی فاصله‌ای را می‌توان به یک نقشه تشبیه کرد که در آن هر واژه یک نقطه و فاصله میان نقاط نشان‌دهنده میزان شباهت معنایی است. در این میان، روابط محلی میان واژه‌ها در مقیاس بزرگتر به یک ساختار منسجم و شبکه‌ای تبدیل می‌شود که می‌توان آن را نقشه معنایی[7] مفاهیم نامید. در این نقشه، واژه‌ها دیگر به‌صورت عناصر منفرد در فرهنگ لغت دیده نمی‌شوند، بلکه در قالب گره‌هایی در یک شبکه معنایی قرار می‌گیرند که

1. Sweet
2. Honey
3. Cream
4. Chocolate
5. Soup
6. Potatoes
7. Semantic Map

براساس حجم عظیم داده‌های متنی شکل گرفته است. درنتیجه، مدل‌های برداری به جای ارائه تعریف‌های ثابت، ساختاری از نزدیکی‌ها، فاصله‌ها و خوشه‌های مفهومی تولید می‌کنند که ماهیتی پویا و وابسته به کاربرد دارند؛ به‌گونه‌ای که با تغییر داده‌های زبانی، ساختار معنایی نیز بازآرایی می‌شود. این ویژگی نشان می‌دهد که معنا در این رویکرد، پدیده‌ای ایستا نیست بلکه در بستر استفاده زبانی شکل می‌گیرد.

با این حال، این نقشه‌های معنایی نیز محدودیت‌هایی دارند؛ از جمله امکان بازتولید سوگیری‌های داده[1] و ناتوانی در درک لایه‌های تفسیری و زمینه‌ای عمیق. در مجموع، گذار از کشف شباهت‌های واژگانی به ترسیم نقشه‌های معنایی، یکی از تحولات بنیادین در فهم محاسباتی زبان است که چشم‌انداز تازه‌ای برای تحلیل متون در علوم انسانی فراهم می‌سازد. بدین ترتیب، فضای برداری بتدریج به یک نقشه معنایی قابل تحلیل تبدیل می‌شود که در آن، روابط میان مفاهیم به‌صورت ساختارمند قابل مشاهده است. با این حال، این نوع بازنمایی همچنان به‌تنهایی قادر نیست سازوکارهای تولید زبان یا فرایند یادگیری معنایی را توضیح دهد. اکنون پرسش این است که آیا می‌توان از همین نمایش‌های برداری، نه فقط برای ترسیم روابط معنایی، بلکه برای ساخت نظام‌هایی استفاده کرد که بتوانند زبان را بیاموزند، تعمیم دهند و تولید کنند؟

### ۵. چگونگی ساخت مدل‌های زبانی بزرگ بر پایه بردارها

ایده بازنمایی برداری، بتدریج به زیربنای نسل جدیدی از مدل‌های زبانی تبدیل شد. این مدل‌ها که امروز با عنوان مدل‌های زبانی بزرگ شناخته می‌شوند و توسط شرکت‌های مطرحی مانند اُپن‌ای‌آی[2] و گوگل[3] توسعه یافته‌اند، در امتداد همان رویکرد بازنمایی برداری معنا عمل می‌کنند. نقطه عطف این مسیر، معماری ترنسفورمر[4] بود. این

1. Data Bias
2. OpenAI
3. Google
4. Transformer

معماری در مقاله‌ای به‌نام «توجه تمام چیزی است که نیاز دارید[1]» توضیح داده شده است. معماری ترنسفورمر نشان داد که می‌توان بدون استفاده از ساختارهای سنتی مانند شبکه‌های عصبی بازگشتی، تنها با استفاده از مکانیزم توجه روابط میان واژه‌ها را در یک جمله مدل‌سازی کرد (واسوانی و همکاران، ۲۰۱۷، صص ۱–۲). در این چارچوب، هر واژه نه یک بردار ثابت، بلکه بازنمایی‌ای وابسته به زمینه[2] دریافت می‌کند؛ به‌گونه‌ای که معنای آن در هر جمله می‌تواند تغییر یابد.

بر این اساس، مدل‌های زبانی بزرگ مانند جی. پی. تی[3] و جِمینای[4] را می‌توان توسعه‌یافته‌ترین صورت‌بندی‌های مدل‌های برداری دانست که در آنها نه‌تنها شباهت واژگانی، بلکه روابط زمینه‌ای و ساختاری میان معانی نیز وارد محاسبات می‌شود. پژوهش‌ها نشان می‌دهند که کارایی این مدل‌ها به‌طور مستقیم به کیفیت بازنمایی‌های برداری آنها وابسته است و همین ساختار فشرده و چندبعدی، امکان تعمیم‌پذیری و پردازش متون گسترده را فراهم می‌سازد (مانینگ، ۲۰۱۵، ص ۷۰۳).

از منظر علوم انسانی، این تحول نشان‌دهنده امکان صورت‌بندی محاسباتی زبان، فرهنگ و معناست. در این چارچوب، مدل‌های زبانی جدید را می‌توان امتداد تکامل‌یافته بازنمایی برداری دانست که در آن معنا از سطح واژه به سطح زمینه و ساختارهای کلان زبانی منتقل شده است.

## روش‌شناسی

روش این پژوهش از نوع مطالعه توصیفی – تحلیلی و با رویکرد مرور و سنتز مفهومی است. گردآوری منابع از طریق جستجوی هدفمند در پایگاه‌های معتبر پژوهشی و با استفاده از کلیدواژه‌های مرتبط با «بازنمایی برداری»، «تعبیه‌سازی کلمات»، «پردازش

1. Attention Is All You Need
2. Contextualized
3. GPT
4. Gemini

زبان طبیعی»، «مدل‌های زبانی بزرگ» و «علوم انسانی دیجیتال» انجام شد. در انتخاب منابع، آثاری مورد توجه قرار گرفتند که از اعتبار علمی برخوردار بوده، نقش بنیادینی در تبیین مفاهیم یا معرفی تحولات این حوزه داشته و ارتباط مستقیمی با موضوع پژوهش داشته باشند. سپس مفاهیم، نظریه‌ها و یافته‌های مرتبط با استفاده از روش تحلیل مفهومی استخراج، مقایسه و در قالب چارچوبی منسجم سازمان‌دهی شدند.

در مرحله تحلیل، تلاش شد ضمن تبیین سیر تحول بازنمایی زبان از رویکردهای اولیه تا مدل‌های زبانی نوین، پیوند میان مبانی فنی بازنمایی برداری و ظرفیت‌های آن در مطالعات علوم انسانی و اسلامی به‌صورت میان‌رشته‌ای بررسی شود. ازاین‌رو، مقاله حاضر بیش از آنکه به ارائه جزئیات فنی الگوریتم‌ها بپردازد، بر تفسیر مفاهیم، تحلیل روندهای علمی و تبیین کاربردهای بازنمایی برداری برای پژوهشگران علوم انسانی و اسلامی تمرکز دارد و می‌کوشد تصویری منسجم از مبانی، تحولات و قابلیت‌های این رویکرد ارائه دهد.

## تحلیل یافته‌ها

### ۱. کاربرد بازنمایی برداری در علوم انسانی

برداری‌سازی کلمات زمانی اهمیت واقعی خود را نشان می‌دهد که از سطح نظری و فنی فراتر رفته و در مسائل واقعی علوم انسانی، علوم اجتماعی و مطالعات فرهنگی به‌کار گرفته شود. این فناوری در اصل ابزاری برای مدل‌سازی زبان است، اما پیامد مهمتر آن، ایجاد امکان تحلیل در مقیاس کلان به‌شمار می‌رود؛ یعنی مطالعه میلیون‌ها متن به‌صورت هم‌زمان و استخراج الگوهایی که در روش‌های سنتی قابل مشاهده نیستند. در این بخش، مهمترین کاربردهای این رویکرد در حوزه علوم انسانی بررسی می‌شود.

نخستین کاربرد مهم، تحلیل تحول معنایی مفاهیم تاریخی است. بسیاری از مفاهیم اجتماعی و فرهنگی در طول زمان دچار تغییر معنا می‌شوند؛ برای مثال، واژه‌هایی مانند «عدالت»، «ملت» یا «آزادی» در دوره‌های مختلف تاریخی معانی متفاوتی داشته‌اند.

مـدل‌های بـرداری ایـن امکـان را فـراهم می‌کننـد کـه مسیر تحـول معنـایی ایـن واژه‌هـا در طول زمـان به‌صـورت کمـی و قابـل مشـاهده تحلیـل شـود. برخـی پژوهش‌هـا نشـان داده‌اند که می‌توان با استفاده از مدل‌های برداری، تغییرات معنایی واژه‌ها را در بازه‌هـای تاریخی مختلف ردیابی کرد و حتـی نقـاط گسسـت معنـایی را شناسـایی کـرد (همیلتون و دیگـران، ۲۰۱۶، ص ۲). ایـن قابلیـت بـرای تاریخ‌نگـاری مفهـومی و مطالعـات اندیشـه بسـیار ارزشمند است.

کاربرد دوم، کشف الگوهای گفتمانی است. در علوم اجتمـاعی، گفتمـان بـه معنـای نظامی از معناها، ارزش‌ها و روایت‌هاست که در زبان بازتاب می‌یابـد. مـدل‌های بـرداری می‌توانند با تحلیـل هم‌رخـدادی واژه‌هـا در حجـم عظیمـی از متـون، خوشـه‌های معنـایی مرتبط با یک گفتمان خاص را شناسایی کنند. بـرای مثـال، در تحلیـل گفتمـان سیاسـی، واژه‌هایی مانند «قدرت»، «حاکمیت»، «امنیت» و «ثبات» معمـولاً در یـک خوشـه معنـایی قرار می‌گیرند. این خوشه‌ها به پژوهشگر کمک می‌کنند ساختارهای فکـری پشـت زبـان سیاسی را شناسایی کند. در پژوهش‌های جدید، نشان داده شده اسـت کـه بردارهـا ابـزار مؤثری برای استخراج ساختارهای گفتمانی پنهان در داده‌های متنی هستند (وورس و کولن، ۲۰۲۰، ص ۲۲۶).

سومین کاربرد مهم، تحلیل کـلان متـون رسـانه‌ای اسـت. در عصـر دیجیتـال، حجـم عظیمی از داده‌های رسانه‌ای روزانه تولید می‌شود که تحلیل دستی آنها عمـلاً غیـرممکن است. بازنمایی برداری امکان تحلیل خودکار این داده‌ها را فراهم می‌کند. برای مثـال در یک پژوهش، متون روزنامـه‌های مختلـف جهـت بررسـی سـوگیری جنسـیت، از طریـق تغییرات معنایی مورد مطالعه قرار گرفته است (وورس و کولن، ۲۰۲۰، ص ۲۳۵). این نوع تحلیل به پژوهشگران رسانه کمک می‌کند تـا الگوهـای بازنمـایی رسـانه‌ای و جهت‌گیری‌هـای گفتمانی را در سطح کلان شناسایی کنند.

چهـارمین کـاربرد، مطالعـه افکـار عمـومی اسـت. شـبکه‌های اجتمـاعی بـه یکـی از مهمترین منابع داده بـرای شـناخت نگرش‌هـای اجتمـاعی تبـدیل شـده‌اند. بـا اسـتفاده از مـدل‌های بـرداری، می‌تـوان احساسـات، نگرش‌هـا و واکنش‌هـای عمـومی نسـبت بـه

موضوعات مختلف را تحلیل کرد. برای مثال، می‌توان بررسی کرد که مردم نسبت به یک سیاست عمومی یا رویداد اجتماعی چه دیدگاه‌هایی دارند و این دیدگاه‌ها چگونه در طول زمان تغییر می‌کنند. در این زمینه، پژوهش‌های متعددی نشان داده‌اند که مدل‌های برداری در ترکیب با تحلیل احساسات[1]، ابزار قدرتمندی برای درک افکار عمومی در مقیاس بزرگ هستند (لیو، ۲۰۱۲، صص ۸–۹).

پنجمین کاربرد، مقایسه و تحلیل روایت خواهد بود (گرگ و دیگران، ۲۰۱۸، ص ۶). فرهنگ‌ها از طریق روایت‌ها شکل می‌گیرند و بازتولید می‌شوند. این روایت‌ها در قالب متون ادبی، رسانه‌ای، تاریخی و دینی ظاهر می‌شوند. مدل‌های برداری می‌توانند ساختارهای روایی مشترک میان متون مختلف را شناسایی کنند و نشان دهند که چگونه مفاهیم فرهنگی در طول زمان بازتولید می‌شوند. برای مثال، می‌توان بررسی کرد که مفهوم «قهرمان» چگونه در متون مختلف فرهنگی بازنمایی شده است و چه ارتباطی میان این بازنمایی‌ها وجود دارد.

ششمین کاربرد مهم، کشف مفاهیم هم‌خانواده در متون بزرگ است. یکی از مشکلات اصلی در تحلیل متون گسترده، شناسایی واژه‌ها و مفاهیمی است که از نظر معنایی به هم نزدیک‌اند اما ممکن است از نظر ظاهری متفاوت باشند. بازنمایی برداری با مدل‌سازی روابط معنایی و زمینه‌ای واژگان، ابزاری موثر در کشف کلمات مترادف و هم‌خانواده فراهم می‌کنند. این روش‌ها با تحلیل شبکه‌های واژگانی و کاربردهای تاریخی کلمات، توانایی شناسایی دقیق‌تر مفاهیم مرتبط و هم‌معنی را ارتقا می‌دهند. (وورس و کولن، ۲۰۲۰، ص ۲۳۲) به این ترتیب، پژوهشگر می‌تواند شبکه‌ای از مفاهیم مرتبط را استخراج کند که در تحلیل سنتی ممکن است نادیده گرفته شوند.

هفتمین کاربرد براساس این رویکرد، تحلیل شبکه مفاهیم قرآنی است. در قرآن کریم، مفاهیم به‌صورت شبکه‌ای از روابط معنایی در کنار یکدیگر قرار گرفته‌اند و درک آنها نیازمند توجه به بافت‌های متعدد است. مدل‌های برداری این امکان را فراهم

1. Sentiment Analysis

می‌کنند که واژگان قرآنی نه به‌صورت منفرد، بلکه در قالب یک شبکه معنایی تحلیل شوند. به این معنا که واژه‌هایی مانند «هدایت»، «نور»، «حق» و «ایمان» می‌توانند در یک فضای معنایی مشترک قرار گیرند و روابط میان آنها به‌صورت ساختاری بررسی شود. این نوع تحلیل می‌تواند به پژوهشگر کمک کند تا الگوهای مفهومی تکرارشونده در متن قرآن را با دقت بیشتری شناسایی کند. بر همین اساس، رفیعی[1] و خدنگی[2]، با بهره‌گیری از مدل‌های یادگیری عمیق همچون آرابرت[3] و ماربرت[4] به تحلیل احساسات و لحن در آیات قرآن کریم پرداخته‌اند (رفیعی و خدنگی، ۲۰۲۵، ص ۱). همچنین در پژوهشی دیگر با عنوان «QSST[5]: ابزار جستجوی معنایی قرآنی»، برداری‌سازی واژگان با آموزش به شیوه سی. بی. اُ. دبلیو[6] بر روی پیکره‌های قرآنی و عربی، مفاهیم را به فضای برداری تبدیل می‌کند. در این روش، تشابه کسینوسی[7] بردار پرسش و موضوعات، آیات مرتبط را بازیابی می‌نماید. نتیجه این پژوهش، بیان‌گر آنست که کاربست بازنمایی برداری متن قرآن، بهبود چشمگیری در جستجوی محتوای آن رقم زده است (محمد و شکری، ۲۰۲۲، ص ۱). این قابلیت، زمینه را برای توسعه سامانه‌های مدیریت دانش اسلامی، موتورهای جست‌وجوی معنایی و تحلیل شبکه مفاهیم دینی فراهم می‌کند. با وجود این تمرکز، در حوزه متون دینی، در حال حاضر شاهد آثار آغازین پژوهشی با زمینه تحلیل برداری متون فارسی هستیم، چنان‌که سرمدی[8] و همکارانش با تمرکز بر بهبود امبدینگ[9] متون فارسی، مدل پیشرفته «حکیم» را معرفی کرده‌اند که بهبود خوبی را نسبت به مدل‌های پیشین خود رقم زده‌است (سرمدی و دیگران، ۲۰۲۵، ص ۱).

1. Rafiei
2. Khadangi
3. AraBERT
4. MARBERT
5. Quranic Semantic Search Tool
6. Continuous Bag-of-Words (CBOW)
یک معماری برداری جهت پیش‌بینی کلمه از روی کلمات همجوار.
7. Cosine Similarity
8. Sarmadi
9. Embedding

### ۲. افق‌های پیشروی علوم اسلامی در تقاطع با بازنمایی برداری

افق‌های کاربردی بازنمایی برداری معنا در مطالعات اسلامی بسیار گسترده است و می‌تواند زمینه‌ساز تحول در شیوه تولید، سازمان‌دهی و بازیابی دانش دینی شود. یکی از مهمترین این کاربردها، کشف روابط معنایی میان احادیث است؛ به گونه‌ای که با استخراج مفاهیم کلیدی و تحلیل روابط میان آنها، امکان شکل‌دهی شبکه‌های معنایی از روایات و دسته‌بندی هوشمند موضوعات حدیثی فراهم می‌شود. این رویکرد همچنین می‌تواند ارتباطات پنهان میان احادیث پراکنده را آشکار سازد؛ قابلیتی که پیش‌تر در مطالعات زبان‌شناسی محاسباتی[1] و پژوهش مبتنی بر بردارهای معنایی میکولوف و همکارانش نیز به اثبات رسیده است. حوزه مهم دیگر، تحلیل تحول معنایی مفاهیم فقهی و کلامی در طول تاریخ است. با استفاده از مدل‌های برداری می‌توان تغییرات معنایی مفاهیمی مانند «اجتهاد»، «عدالت»، «بیع» و سایر اصطلاحات کلیدی را در دوره‌های مختلف ردیابی کرد و سیر تحول فهم و تفسیر آنها را مورد مطالعه قرار داد؛ رویکردی که کارآمدی آن در مطالعات زبان‌شناسی تاریخی نیز تأیید شده است (همیلتون و دیگران، ۲۰۱۶، ص ۶). افزون بر این، بازنمایی برداری می‌تواند زیرساخت توسعه موتورهای جست‌وجوی معنایی در متون اسلامی را فراهم آورد؛ به نحوی که بازیابی اطلاعات نه براساس تطابق لفظی، بلکه بر پایه نزدیکی معنایی مفاهیم انجام شود و پژوهشگران بتوانند حتی بدون استفاده از عبارات دقیق موجود در متن، به منابع مرتبط دسترسی یابند. پژوهش‌ها حاکی از آنست که استفاده از روش‌های مذکور قدرت خوبی را در جستجوی معنایی در اختیار پژوهشگران قرار می‌دهد (میرعرب، ۱۴۰۲، ص ۲۰۱). در سطحی کلان‌تر، این فناوری ظرفیت آن را دارد که به ابزاری برای مدیریت دانش اسلامی تبدیل شود؛ بدین معنا که میراث عظیم متون دینی در قالب شبکه‌ای یکپارچه از مفاهیم، موضوعات و روابط معرفتی سازمان‌دهی شود و امکان فهم نظام‌مندتر پیوندهای میان دانش‌های فقهی، تفسیری، حدیثی، اخلاقی و کلامی را فراهم سازد. از این منظر،

1. Computational Linguistics

بازنمایی برداری معنا نه تنها یک ابزار فنی در پردازش زبان، بلکه بستری نوین برای توسعه مطالعات اسلامی هوشمند و گشودن روزنه‌های تازه در پژوهش‌های دینی به شمار می‌آید.

### ۳. محدودیت‌ها و سوءفهم‌ها

با وجود تمام پیشرفت‌های چشمگیر در بازنمایی برداری و مدل‌های زبانی بزرگ، ضروری است که نگاه انتقادی به این فناوری حفظ شود؛ بویژه در علوم انسانی و اسلامی که با مفاهیم تفسیری، تاریخی و ارزشی سر و کار دارند.

یکی از مهمترین محدودیت‌های این مدل‌ها، مسئله سوگیری داده‌ها است. از آنجاکه مدل‌های برداری براساس حجم عظیمی از داده‌های متنی آموزش می‌بینند، هرگونه سوگیری فرهنگی، اجتماعی یا تاریخی موجود در این داده‌ها در ساختار مدل نیز بازتولید می‌شود. به عبارت دیگر، اگر داده‌ها دارای نگاه‌های خاص فرهنگی یا جنسیتی باشند، مدل نیز همان الگوها را یاد می‌گیرد و تقویت می‌کند. استفاده از فناوری هوش مصنوعی در امتداد تعصب یا بی‌عدالتی، به‌عنوان یکی از چالش‌های جدی هوش مصنوعی مطرح شده است (کالیسکان و دیگران، ۲۰۱۷، ص ۱).

محدودیت دیگر، ناتوانی در درک زمینه تاریخی و فرهنگی است. مدل‌های برداری و حتی مدل‌های زبانی بزرگ، متن را عمدتاً به‌عنوان یک داده مستقل از زمان و زمینه تحلیل می‌کنند. در حالی که در علوم انسانی، معنا همواره وابسته به زمینه تاریخی، اجتماعی و فرهنگی است. برای مثال، واژه‌ای مانند «آزادی» در متون قرن نوزدهم با معنای امروزی آن تفاوت‌های جدی دارد، اما مدل‌های ماشینی ممکن است این تفاوت را به‌درستی تشخیص ندهند، مگر اینکه به‌صورت خاص آموزش داده شوند. این محدودیت نشان می‌دهد که فهم تاریخی، همچنان یک چالش بنیادین برای مدل‌های محاسباتی است.

از سوی دیگر، مسئله استعاره، کنایه و لایه‌های تفسیری نیز یکی از نقاط ضعف مهم این مدل‌هاست. زبان انسانی سرشار از بیان‌های غیرمستقیم است که معنا را نه در سطح

واژه، بلکه در سطح زمینه و فرهنگ منتقل می‌کنند. برای مثال، جمله «دستش باز است» در بسیاری از موارد معنایی کاملاً غیرمادی دارد، اما مدل‌های برداری ممکن است آن را صرفاً براساس هم‌رخدادی واژه‌ها تحلیل کنند. مطالعات پژوهشی اخیر بیان‌گر آنست که توانایی فهم استعاره‌ها توسط مدل‌های هوش مصنوعی یکی از ابهامات پژوهشگران حوزه علوم شناختی و زبان‌شناسی محاسباتی است (میچل و کراکاور، ۲۰۲۳، ص ۲).

در کنار این محدودیت‌ها، یک سوءفهم مهم دیگر نیز وجود دارد: تصور اینکه افزایش حجم داده و قدرت محاسباتی به‌تنهایی می‌تواند منجر به فهم انسانی شود. این دیدگاه که با نوعی از خوش‌بینی به فناوری همراه است، فرض می‌کند که پیچیدگی آماری در نهایت به فهم منجر خواهد شد (میچل و کراکاور، ۲۰۲۳، ص ۲). در طرف مقابل، برخی معتقدند که ادعای خوش‌بینی فناورانه نادرست است، زیرا درک انسانی فقط با افزایش مقیاس داده و پارامترها حاصل نمی‌شود؛ بلکه درک حقیقی مستلزم عواملی بنیادی‌تر همچون تجربه زیسته، آگاهی و زمینه فرهنگی است. از این منظر، مدل‌های زبانی بزرگ با وجود تولید زبانی روان، فاقد چنین مؤلفه‌هایی هستند و آنچه آموخته‌اند صرفاً الگوهای صوری زبان است نه معنای مبتنی بر جهان خارج (میچل و کراکاور، ۲۰۲۳، صص ۲–۳).

در فضای علوم انسانی، این محدودیت‌ها اهمیت دوچندان دارند. زیرا خطر آن وجود دارد که ابزارهای محاسباتی بجای ابزار تحلیل، به منبع تفسیر نهایی تبدیل شوند. در حالی‌که در سنت علوم انسانی، تفسیر همواره یک فرآیند انسانی، انتقادی و زمینه‌مند بوده است. بنابراین، استفاده از مدل‌های برداری باید همواره در چارچوبی انتقادی و مکمل صورت گیرد، نه جایگزین روش‌های تفسیری.

در نهایت، باید تأکید کرد که ارزش اصلی بازنمایی برداری، نه در جایگزینی روش‌های موجود بلکه در توانایی آن برای افزودن یک لایه تحلیلی جدید به علوم انسانی است. این فناوری می‌تواند الگوهای پنهان را آشکار کند، اما تفسیر این الگوها همچنان نیازمند دانش انسانی، زمینه‌شناسی و تحلیل انتقادی است. به همین دلیل، رویکرد صحیح در استفاده از این ابزارها، رویکردی ترکیبی است که در آن محاسبات ماشینی و فهم انسانی در کنار یکدیگر قرار می‌گیرند.

### ۴. آینده علوم انسانی در عصر معناهای محاسباتی

تحولاتی که در این مقاله بررسی شد، نشان می‌دهد که علوم انسانی بتدریج وارد مرحله‌ای تازه می‌شود که در آن، تحلیل متون در مقیاس کلان و کار با ابزارهای محاسباتی به بخشی از روش پژوهش بدل خواهد شد. در این وضعیت، نقش پژوهشگر نیز دگرگون می‌شود؛ او دیگر فقط مفسر متن نخواهد بود، بلکه باید بتواند میان فهم تفسیری و تحلیل الگوریتمی پیوند برقرار کند. مطالعات جدید در حوزه علوم انسانی دیجیتال نیز نشان می‌دهند که داده‌کاوی و سازمان‌دهی دیجیتال منابع تاریخی، افق‌های تازه‌ای برای مطالعه تاریخ اندیشه و تحلیل روابط پنهان در متون می‌گشاید (رحمان، ۲۰۲۵، ص ۲). از این منظر، آینده علوم انسانی نه در حذف تفسیر انسانی، بلکه در هم‌نشینی آن با ابزارهای محاسباتی و گسترش ظرفیت‌های روش‌شناختی آن رقم خواهد خورد (گرایمر و استوارت، ۲۰۱۳، صص ۲۸–۲۹).

## نتیجه‌گیری

بازنمایی برداری معنا را می‌توان یکی از تحولات مهم پردازش زبان طبیعی در دهه‌های اخیر دانست. این رویکرد با تبدیل واژه‌ها و متون به فضاهای برداری چندبعدی، امکان مدل‌سازی روابط معنایی، سنجش شباهت مفاهیم و تحلیل ساختارهای زبانی را در مقیاسی فراهم کرده است که پیش‌تر با روش‌های سنتی قابل دستیابی نبود. بررسی سیر تحول بازنمایی زبان نیز نشان داد که بسیاری از معماری‌های نوین هوش مصنوعی، از مدل‌های تعبیه‌سازی کلمات تا مدل‌های زبانی بزرگ، بر همین بنیان استوار شده‌اند.

یافته‌های این مقاله نشان می‌دهد که بازنمایی برداری می‌تواند ظرفیت‌های قابل توجهی برای مطالعات علوم انسانی و اسلامی فراهم آورد. از جمله این ظرفیت‌ها می‌توان به تحلیل تحول تاریخی مفاهیم، کشف روابط معنایی میان متون، توسعه جست‌وجوی معنایی و پشتیبانی از سامانه‌های مدیریت دانش اشاره کرد. با این حال، این فناوری جایگزین فهم و تفسیر انسانی نیست. مدل‌های برداری معنا را به‌صورت محاسباتی بازنمایی می‌کنند، اما فاقد تجربه زیسته، قصد و آگاهی انسانی هستند.

ازاین‌رو، ارزش اصلی آنها نه در جانشینی با پژوهشگر، بلکه در تقویت توان او برای تحلیل حجم گسترده‌ای از داده‌های متنی و آشکارسازی الگوهایی است که در روش‌های متعارف کمتر قابل مشاهده‌اند.

بر این اساس، بهره‌گیری از بازنمایی برداری را باید گامی در جهت توسعه روش‌های پژوهش در علوم انسانی و اسلامی دانست؛ رویکردی که می‌تواند در کنار سنت‌های تفسیری و تحلیلی موجود، افق‌های تازه‌ای برای مطالعه نظام‌مند معنا در متون و میراث فکری فراهم سازد.

# فهرست منابع


میرعرب، علی. (۱۴۰۲). بررسی راهکارهای جستجو و بازیابی معنایی متون فارسی و عربی. علوم و فنون، ۹(۴)، صص ۱۸۵–۲۰۴.

Almeida, Felipe; & Xexéo, Geraldo. (2019). Word Embeddings: A Survey. CoRR, abs/1901. 09069. https://doi. org/10. 48550/arXiv. 1901. 09069

Caliskan, Aylin; Bryson, Joanna J. ; & Narayanan, Arvind. (2017). Semantics Derived Automatically from Language Corpora Contain Human-Like Biases. Science, 356(6334), 183–186. https://doi. org/10. 1126/science. aal4230

Ferrone, Lorenzo; & Zanzotto, Fabio Massimo. (2020). Symbolic, Distributed, and Distributional Representations for Natural Language Processing in the Era of Deep Learning: A Survey. Frontiers in Robotics and AI, 6, 153. https://doi. org/10. 3389/frobt. 2019. 00153

Garg, Nikhil; Schiebinger, Londa; Jurafsky, Dan; & Zou, James. (2018). Word Embeddings Quantify 100 Years of Gender and Ethnic Stereotypes. Proceedings of the National Academy of Sciences, 115(16), E3635–E3644. https://doi. org/10. 1073/pnas. 1720347115

Grimmer, Justin; & Stewart, Brandon M. (2013). Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. Political Analysis, 21(1), 1–31. https://doi. org/10. 1093/pan/mps028

Hamilton, William L. ; Leskovec, Jure; & Jurafsky, Dan. (2016). Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In K. Erk & N. A. Smith (Eds.), (K. Erk & N. A. Smith, Eds.), Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers) (pp. 1489–1501). Berlin, Germany: Association for Computational Linguistics. https://doi. org/10. 18653/v1/P16-1141

Jurafsky, Daniel; & Martin, James H. (2026). Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models (3rd ed.) .

Liu, Bing. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers .

Manning, Christopher D. (2015). Computational Linguistics and Deep Learning. Computational Linguistics, 41(4), 701–707. https://doi. org/10. 1162/COLI_ a_00239

Merrill, William; Goldberg, Yoav; Schwartz, Roy; & Smith, Noah A. (2021). Provable Limitations of Acquiring Meaning from Ungrounded Form: What Will Future Language Models Understand? Transactions of the Association for Computational Linguistics, 9, 1047–1060. https://doi. org/10. 1162/tacl_a_00412

Mikolov, Tomas; Chen, Kai; Corrado, Greg; & Dean, Jeffrey. (2013). Efficient Estimation of Word Representations in Vector Space. arXiv:1301. 3781 [Cs:CL]. https://doi. org/10. 48550/arXiv. 1301. 3781

Mitchell, Melanie; & Krakauer, David C. (2023). The Debate Over Understanding in AI's Large Language Models. Proceedings of the National Academy of Sciences, 120(13), e2215907120. https://doi. org/10. 1073/pnas. 2215907120

Mohamed, E. H. ; & Shokry, E. M. (2022). QSST: A Quranic Semantic Search Tool Based on Word Embedding. Journal of King Saud University: Computer and Information Sciences, 34(3), 934–945. https://doi. org/10. 1016/j. jksuci. 2020. 01. 004

Rafiei, Mahdi; & Khadangi, Ehsan. (2025). Sentiment and Tone Analysis of the Holy Qur'an Using Natural Language Processing. Journal of Interdisciplinary Qur'anic Studies, 4(2). https://doi. org/10. 37264/JIQS. V4I2. 5

Rahman, A. M. (2025). Digital Humanities: Computational Methods for Research. International Research Journal of Arts and Social Sciences, 13(2), 1–5. https://doi. org/http:/dx. doi. org/10. 14303/2276-6502. 2025. 114

Sarmadi, Mehran; Alikhani, Morteza; Zinvandi, Erfan; & Pourbahman, Zahra. (2025). Hakim: Farsi Text Embedding Model. arXiv:2505. 08435 [Cs:CL, Cs:Cs:AI, Cs:Cs:LG]. https://doi. org/DOI:10. 48550/arXiv. 2505. 08435

Schomacker, Thorben; & Tropmann-Frick, Marina. (2021). Language Representation Models: An Overview. Entropy, 23(11), 1422. https://doi. org/10. 3390/e23111422

Vaswani, Ashish; Shazeer, Noam; Parmar, Niki; Uszkoreit, Jakob; Jones, Llion; Gomez, Aidan N. ; … Polosukhin, Illia. (2017). Attention Is All You Need. CoRR, abs/1706. 03762. https://doi. org/10. 48550/arXiv. 1706. 03762

Wevers, Melvin; & Koolen, Marijn. (2020). Digital Begriffsgeschichte: Tracing Semantic Change Using Word Embeddings. Historical Methods, 53(4), 226–243. https://doi. org/10. 1080/01615440. 2020. 1760157

Wohlgenannt, Gerhard; Chernyak, Ekaterina; Ilvovsky, Dmitry I. ; Barinova, Ariadna; & Mouromtsev, Dmitry. (2019). Relation Extraction Datasets in the Digital Humanities Domain and their Evaluation with Word Embeddings. CoRR, abs/1903. 01284. https://doi. org/10. 48550/arXiv. 1903. 01284