



# 4<sup>th</sup> International Conference on Data Mining, Big Data and Machine Learning (DBML 2026)

March 30 ~ 31, 2026, Virtual Conference

## **Call for Participation**

We invite you to join us in **4<sup>th</sup> International Conference on Data Mining, Big Data and Machine Learning (DBML 2026)**

This conference will serve as a premier international forum for researchers, practitioners, and industry experts to exchange knowledge, discuss breakthroughs, and explore emerging trends in the rapidly evolving fields of Data Mining, Big Data, and Machine Learning. As data continues to grow in scale, complexity, and impact, DBML 2026 provides a vital platform for presenting innovative theories, advanced methodologies, and practical applications that address the challenges and opportunities of data driven intelligence.

## **Highlights of DBML 2026 include:**

- [4<sup>th</sup> International Conference IOT, Blockchain and Cryptography \(IOTBC 2026\)](#)
- [8<sup>th</sup> International Conference on Bioscience & Engineering \(BIO 2026\)](#)
- [4<sup>th</sup> International Conference on Education in Post Pandemic \(EDUPAN 2026\)](#)
- [4<sup>th</sup> International Conference on Computing and Information Technology \(CITE 2026\)](#)
- [4<sup>th</sup> International Conference on NLP & AI \(NLPAI 2026\)](#)
- [4<sup>th</sup> International conference on Health Informatics \(HEIN 2026\)](#)
- [8<sup>th</sup> International Conference on Mechanical Engineering \(MECH 2026\)](#)
- [9<sup>th</sup> International Conference on Electrical and Electronics Engineering \(ICEEE 2026\)](#)
- [12<sup>th</sup> International Conference on Emerging Trends in Electrical, Electronics & Instrumentation Engineering \(EEI 2026\)](#)

## **Registration Participants**

Non-Author / Co-Author/ Simple Participants (no paper)

**100 USD (With proceedings)**

Here's where you can reach us : [dbml@dbml2026.org](mailto:dbml@dbml2026.org) or [dbmlconference70@gmail.com](mailto:dbmlconference70@gmail.com)

## **ACCEPTED PAPERS**

**Artificial Intelligence and NLP on Reddit: Unsupervised Detection of Food Trends and Healthy Eating Patterns**

Rocío del Campo-Pedrosa<sup>1</sup>, Diego del Campo-Pedrosa<sup>1</sup>, Bettina Merlin<sup>2</sup> and Ana González-Marcos<sup>1</sup>,  
1Department of Mechanical Engineering, Universidad de La Rioja, Logroño, La Rioja, Spain, 2Fakultät  
International Business, Hochschule Heilbronn, Heilbronn, Germany

## **ABSTRACT**

Traditional sensory analysis in food innovation provides limited insight into consumer behavior, whereas social platforms such as Reddit offer large-scale, real-time textual data on food-related practices and perceptions. This study evaluates Reddit as a scalable source for detecting food trends and healthy eating patterns in Spanish-language discussions using artificial intelligence (AI) and natural language processing (NLP). An end-to-end pipeline was implemented, including targeted data scraping across seven food-related domains, Spanish-language filtering ( $\geq 70\%$  confidence), customized preprocessing, and unsupervised topic discovery via k-means clustering. The system processed 17,774 Spanish-language posts from an initial corpus of 92,949 entries. Despite linguistic challenges such as polysemy and lemmatization errors, the method produced coherent and representative themes, including barriers to home cooking, weight management concerns, economic factors, food categories, and nutrition-related consultations. These results demonstrate the effectiveness of unsupervised NLP techniques for large-scale monitoring of food-related discourse on social media.

## **Keywords**

Natural Language Processing, Unsupervised Learning, Social Media Mining, Artificial Intelligence.

## **A Methodological Approach to Calligraphic Obfuscation**

Bettina Merlin and Ana González-Marcos, Department of Computer Science, College of Computer, Qassim University, Buraydah, 51452, Saudi Arabia

## **ABSTRACT**

As automated optical character recognition (OCR) and deep learning-based solvers achieve near-human accuracy in breaking conventional CAPTCHAs, there is a critical need for security mechanisms that exploit the inherent limitations of machine perception. This paper proposes a novel methodological framework for "Calligraphic Obfuscation," a security-by-design approach that leverages the structural complexity and fluid entropy of traditional Arabic calligraphic styles. Unlike standard text-based challenges, our approach introduces a multi-phase generation pipeline that systematically maps linguistic strings into high-complexity visual domains. The methodology integrates a four-tier classification of calligraphic fonts—ranging from high-legibility styles like Naskh to high-entropy scripts such as Shakstah—and augments them with an adversarial layer utilizing Jacobian-based Saliency Map Attacks (JSMA). By formalizing the transition from cloud-centric generation to resource-efficient on-device architectures, this study provides a repeatable blueprint for developing robust, human-interactive proofs. The proposed framework offers a dual-benefit: significantly increasing the computational cost for adversarial machine learning models while maintaining a sustainable cognitive load for human users. This work lays the foundation for a new generation of linguistically-diverse and adversarially-hardened authentication challenges tailored for modern, resource-constrained mobile environments.

## **Keywords**

Calligraphic Obfuscation, CAPTCHA Security, Adversarial Machine Learning, Arabic Script Complexity, Human-Interactive Proofs, JSMA.

## **Δ.72 Wearable Exam Stress Validation Report: Field Equation Verification and Biological Coherence Analysis**

Allison Hensgen, Independent Researcher, USA

### **ABSTRACT**

This report presents an empirical validation of the  $\Delta.72$  field equation within a real-world physiological dataset measuring stress responses in students during examination periods. The  $\Delta.72$  model posits that coherence within biological systems arises from dynamic alignment between internal variability, environmental fields, and phase-synchronized information flow. Using the publicly available Wearable Exam Stress dataset, this study tests whether measurable alignment (A), phase coherence ( $\lambda$ ), and emergent output (E) follow the theoretical relationships predicted by the  $\Delta.72$  equation. Results show statistically significant coupling between field alignment and emergent coherence, supporting the model's claim that adaptive, not rigid, synchrony underlies systemic health.

## **Evaluating AI-readiness of Unstructured Data in Organization: A Lightweight Automated Quality Scoring Framework for Generative AI Adoption**

Veer Bobade, Sujal Zade, Prathmesh Waghmare, Sumit Ladwan, Karan Bhute, Sohan Akare, DMIHER University, India

### **ABSTRACT**

Generative Artificial Intelligence (GenAI) has become the key issue of modern organizations. The pre-deployment test is very important to ensure that the unstructured data used by such systems is strong enough to produce reliable and consistent results. Structured data are therefore advantageous in that they have the well-known quality control mechanisms in place, as well as governance systems, whereas unstructured data, which constitute approximately 80 per cent of organizational information, are often out of concern, disorganized, and potentially dangerous. Poorly trained documentation may cause model hallucinations, in which the AI does not make the correct inferences or have poor results, and thus undermines user trust in GenAI deployments. In order to address this problem, we introduce Lightweight Automated Quality Scoring Framework (LAQS) a framework which can be used to assess the preparedness of unstructured data to be utilized by AI. LAQS adopts a single paradigm of scoring, which is a combination of linguistic quality, semantic cohesion, structural consistency, information extractability. It focuses on five major dimensions, including completeness (i.e. the inclusion of the necessary content), consistency (a regular formatting), clarity (readable by humans), semantic coherence (logical flow), and extractability (readable by machine). These dimensions are operationalized through a systematic and step-based process which may produce credible scores in readiness. In the validated work on the LAQS, we tested it on the CORD-19 corpus with 1,000 academic articles. There was substantial improvement in the presentation of documents with high scores in readiness, semantic similarity improved by 6570, summary accuracy improved by 4050, and the frequency of hallucination decreased up to four times. These findings indicate the extreme importance of unstructured data quality in GenAI achievement, and these arguments can dispel the idea that this kind of quality can be handled after the event. Based on this, we propose to introduce such tools like LAQS into the structures of AI programmes, data-engineering to be a part of AI governance to enable ongoing assessment.

### **Keywords**

Unstructured Data Quality Assessment, Readiness for Generative Ai, Governance in Data Management, Reducing Hallucinations, Maintaining Semantic Coherence.

### **Compliance with Data Privacy and Confidentiality Protocols in Ghanaian Health Facilities: A Health Systems Analysis**

Lily Obenewaa Asare<sup>1,2</sup>, James Antwi<sup>1,3</sup>, <sup>1</sup>Ghana Adventist Health Services, Ghana, <sup>2</sup>University of Hull, Hull, United Kingdom, <sup>3</sup>Centre for Health and Social Policy Research, Ghana

#### **ABSTRACT**

This qualitative study examines how health system structures in Ghana impact nursing and midwifery adherence to patient data privacy and confidentiality protocols. The research involved researcher observations and focus group discussions with 60 healthcare professionals across six hospitals, using a clinical vignette for thematic analysis of practical situations. Findings demonstrate weaknesses across key health system building blocks that negatively impact patient information privacy and confidentiality. The findings indicate critical shortcomings within the leadership and governance, health workforce, health information management, and service delivery domains, emphasizing the need for comprehensive, holistic system strengthening to ensure robust privacy protection. As adaptive and complex entities, health systems are fundamental to safeguarding patients' medical information. However, operating these systems in silos undermines the efficacy of privacy protections within clinical environments. Policy and practice frameworks need to adopt a holistic approach to systems strengthening to ensure robust protection of patient data.

#### **Keywords**

Privacy, Confidentiality, Health Systems, Patients Information.

### **Reimagining The Teacher: A Design-Based Inquiry Into Mentoring, Values-Pedagogy, and GCED Within NEP 2020**

Ashok K. Pandey, Global Peace Foundation, USA

#### **ABSTRACT**

The National Education Policy 2020 places the teacher at the heart of educational transformation, yet the pathway from policy aspiration to lived classroom practice remains insufficiently explored. Mentoring is a formative space where such transformation can occur meaningfully. The present study explores how Panchkosha-aligned mentoring cultivates reflective capacity, ethical awareness, and values-led pedagogy among teachers, using a design-based practitioner inquiry. Evidence was collected through mentee reflections, continuous professional development logs, and an institutional case vignette involving 50 teachers and school leaders across mentoring cohorts between 2023 and 2025. The findings support a shift in teacher identity from an instructor to a mentor. Emotional alignment emerges as a prerequisite for learning. Ethical discernment emerges as a marker of professional maturity, and joy as a legitimate indicator of growth. The values such as empathy, cooperation, peace, and responsibility emerge organically through reflective practice. These values are associated with Global Citizenship Education (GCED). The findings of the study offer a globally intelligible model of teacher development by integrating Indian Knowledge Systems with contemporary educational discourse. The paper argues that

mentoring serves as a crucial bridge between NEP 20's structural reforms and the everyday realities of teaching and learning when grounded in inner awareness and sustained dialogue.

### **Keywords**

Teacher Mentoring, Panchkosha Framework, NEP 2020, Global Citizenship Education, Reflective Practice.

## **Privacy Preserving Techniques for Patient Medical Records Using Blockchain Technology**

Preet Parekh, Patel Chhayank Kaushikbhai and Dr. Hemamalini V, SRM Institute of Science and Technology, Chennai, India

### **ABSTRACT**

In contemporary health care, the greatest issues are patient data security and privacy, with respect to increasing digitization. Traditional methods of dealing with such data have vulnerabilities such as incorrect use, violations, and incompatibilities.

### **Keywords**

Blockchain, Zero-Knowledge Proofs (ZKPs), Tamper-Proof Storage, Decentralized Systems, Privacy-Preserving Verification, Data Interoperability, Regulatory Compliance, HIPAA, GDPR, Secure Data Exchange.

## **Beyond Devices: Digital Equity as the Foundation of K–12 Emergency Preparedness**

Joseph Kenny Vermeille, Doctor Vermeille Global EdTech & Research LLC, USA

### **ABSTRACT**

The COVID-19 pandemic revealed that the digital divide is not merely an equity challenge but a foundational emergency preparedness failure. This paper introduces the Digital Equity as Emergency Infrastructure (DEEI) framework, organized around four interdependent pillars: device equity, connectivity equity, digital literacy equity, and AI and multilingual literacy equity. Drawing on nine dimensions of digital access barriers as distinct failure modes, U.S. national data, and global case studies from conflict zones (Gaza, Ukraine, Haiti, Sudan), this paper argues that digital equity—including AI literacy, multilingual access, disability access, and offline contingency—is a structural component of K–12 crisis preparedness. Updated through the DoctorVermeille EdTech Framework [7] and its companion journal article [6], the paper proposes that instructional design is a moral practice: emergency plans that ignore digital equity are not neutral—they are complicit in the inequities they amplify. Multi-level policy recommendations are provided.

### **Keywords**

Digital Equity, Emergency Preparedness, K–12 Education, Deei Framework, Digital Divide.

# Higher Education 4.0 and the Future of Digital Learning Environments: Platforms, Openness and Scalable Learning for Sustainable Transformation

Vera Alpar, Metropolitan University, Hungary

## ABSTRACT

Digital learning environments increasingly determine how universities deliver quality at scale while pursuing equitable learning outcomes and sustainable transformation. This paper consolidates recent effectiveness studies to show where next generation platforms deliver measurable gains and where results depend on design and context. Across 21 empirical evaluations summarized in a 2015–2025 systematic review, AI-driven tools and adaptive systems are associated with roughly 15–35 percent improvements in course performance and higher task efficiency and satisfaction on average, with effect heterogeneity across disciplines and implementations (Kwak, 2025). Additional syntheses of adaptive learning platforms report significant gains in academic performance and motivation particularly in STEM fields, conditioned by infrastructure and instructor training (Núñez-Hernández, Avilés-Castillo, and Buele, 2025). Randomized and quasi-experimental studies illuminate boundary conditions: AI instructional agents can raise perceived learner control, interaction frequency, and post-test scores relative to MOOC plus chatbot and lecture formats in controlled settings (Qin et al., 2025), while stand-alone chatbots may yield no significant outcome advantage without aligned pedagogy and assessment (Eteng-Uket and Ezeoguine, 2025). At scale, adaptive and intelligent MOOCs reduce dropout and increase completion compared to non-adaptive designs, and automated formative feedback materially shifts learner progress (Papadimitriou, 2023; Adaptive Learning at Scale, 2024). The analysis translates this evidence into design and leadership guidance for Higher Education 4.0, aligning platforms, assessment, analytics, open resources, and partnerships with the Sustainable Development Goals (Zou et al., 2025; Graham et al., 2023).

## Keywords

Education 4.0, Higher Education 4.0, Digital Learning Environments, Adaptive Learning, Ai Instructional Agents, Chatbots, Moocs, Oer, Learning Analytics, SDGS, Governance Equity, Innovative Policy.

## Penrec: A Recommender System for Pension Savings

Onaopepo Adekunle<sup>1</sup>, Chang Sun<sup>1</sup>, Arno Ried<sup>1</sup>, Michel Dumontier<sup>1</sup>, <sup>1</sup>Maastricht University, Netherlands, <sup>2</sup>University of Hull, Hull, United Kingdom, <sup>3</sup>Centre for Health and Social Policy Research, Ghana

## ABSTRACT

Recommender systems are ubiquitous in various domains ranging from e-commerce to financial services to deliver personalized products and services at scale. This work designs and develops a proof of concept for personalized pension recommendations. PenRec, a novel pension savings recommender system, takes the combination of individuals' demographic data and risk preferences data as input. PenRec combines collaborative filtering and deep learning methods to provide individual savings recommendations based on individuals with similar characteristics as pension savings advice. To estimate the risk preferences of individuals, we implemented a survey, including incentivized economic experiments, among the Dutch working population (N = 4, 282); upon which we employed supervised learning techniques to estimate the risk preferences of the general population of the Netherlands based on register data from Statistics Netherlands. We present an analysis of the comparison between traditional collaborative filtering and

deep learning for pension savings recommendations. Additionally, we investigate whether the inclusion of a behavioral measure as a feature, estimated risk preference in particular together with demographic data is important for pension savings recommendation. We find that traditional collaborative filtering tends to recommend average values of individual savings for specified categories of people such as self-employed, income source, and province with no deviations for such categories; while the deep learning-based recommender system can capture varying categories of people in the group of similar individuals and hence provides recommendations that track the real savings more closely relative to the traditional collaborative filtering method. In addition, we observe the inclusion of risk preference variable as a feature with which to compute individual similarity impacts the recommended savings marginally and encourages the addition of behavioral measures for a robust recommendation.

## **Keywords**

Recommender System, Pensions, Collaborative filtering, Deep Learning.

## **PENREC: A RECOMMENDER SYSTEM FOR PENSION SAVINGS**

Onaopepo Adekunle<sup>1</sup>, Chang Sun<sup>1</sup>, Arno Riedl<sup>2</sup>, Michel Dumontier<sup>1</sup>

<sup>1</sup>Department of Advanced Computing Sciences, Maastricht University, Netherlands

<sup>2</sup>Behavioral Economics Department, Maastricht University, Netherlands

## **ABSTRACT**

Recommender systems are ubiquitous in various domains ranging from e-commerce to financial services to deliver personalized products and services at scale. This work designs and develops a proof of concept for personalized pension recommendations. PenRec, a novel pension savings recommender system, takes the combination of individuals' demographic data and risk preferences data as input. PenRec combines collaborative filtering and deep learning methods to provide individual savings recommendations based on individuals with similar characteristics as pension savings advice. To estimate the risk preferences of individuals, we implemented a survey, including incentivized economic experiments, among the Dutch working population (N = 4, 282); upon which we employed supervised learning techniques to estimate the risk preferences of the general population of the Netherlands based on register data from Statistics Netherlands. We present an analysis of the comparison between traditional collaborative filtering and deep learning for pension savings recommendations. Additionally, we investigate whether the inclusion of a behavioral measure as a feature, estimated risk preference in particular together with demographic data is important for pension savings recommendation. We find that traditional collaborative filtering tends to recommend average values of individual savings for specified categories of people such as self-employed, income source, and province with no deviations for such categories; while the deep learning-based recommender system can capture varying categories of people in the group of similar individuals and hence provides recommendations that track the real savings more closely relative to the traditional collaborative filtering method. In addition, we observe the inclusion of risk preference variable as a feature with which to compute individual similarity impacts the recommended savings marginally and encourages the addition of behavioral measures for a robust recommendation.

## **.KEYWORDS**

Recommender System, Pensions, Collaborative filtering, Deep Learning