Driving media innovation through collaborative artificial intelligence

Received (in revised form): 17th September, 2024

Alexandre Rouxel

Senior Project Manager, Technology and Innovation, European Broadcasting Union, Switzerland

Alexandre Rouxel is a senior project manager in the Technology and Innovation of the European Broadcasting Union (EBU). In this role, he coordinates collaborative developments in artificial intelligence and data with EBU members, and leads the development of the AI Hub and algorithms for media content analysis, tagging and enrichment. In international projects, he serves as the task leader responsible for specifying the ontology to exchange media content in the Trusted European Media Data Space. He also contributes to standards development through SMPTE, with a focus on AI and metadata. Before joining the EBU, he spent 20 years working as an algorithm and systems engineer at Nasdaq-listed companies, where he designed algorithms for products and contributed to industry standards.

Department of Technology and Innovation, European Broadcasting Union, L'Ancienne-Route 17A, 1218 Le Grand-Saconnex, Switzerland
E-mail: rouxel@ebu.ch

Alberto Messina

Principal Applied Scientist — Media and AI Research, RAI, Italy

Alberto Messina leads research into the application of artificial intelligence at RAI's Research and Development Unit. In this role, he coordinates research and development and innovation activities in various fields including archive information management, automated production, intelligent unmanned aerial vehicle systems, advanced digital accessibility and publishing services. A former professor at the Polytechnic of Turin, Alberto holds a degree in electronic engineering and a PhD in business and management with specialisation in computer science, and is the author of more than 100 scientific publications. He has worked on numerous projects funded by the European Commission, and is an active participant in multiple international standardisation bodies, including MPEG and the European Broadcasting Union.

Media & AI, Centro Ricerche, Innovazione Tecnologica e Sperimentazione, Via Cavalli, 6, 10138 Torino, Italy
E-mail: alberto.messina@rai.it

Ivan Thomas

Head of Research and Development, Radio France, France

Ivan Thomas is Head of Research and Development within the Digital Department of Radio France. With a strong background in digital media and technology innovation, Ivan leads a team dedicated to advancing Radio France's digital capabilities. Their work focuses on developing cutting-edge solutions to enhance user experience, like the semi-automatic segmentation of podcasts or the detection of highlights.

Radio France, Service DNSI, 116 avenue du Président Kennedy, 75016 Paris, France
E-mail: ivan.thomas@radiofrance.com

Tatjana Mladenovic

Lead Data Manager, BBC Archive Technology & Services, UK

Tatjana Mladenovic is a lead data manager at the BBC's Archive Technology & Services department. She also chairs the EBU AI and Metadata group. She has more than 20 years of experience in the field of metadata for media, having held roles in engineering, data analytics and management. Her professional interests include the automation of data processes, data modelling and data quality.

Archive Technology & Services, Zone 2.27, BBC, 40 Pacific Quay, Glasgow G51 1DA, UK
E-mail: tatjana.mladenovic@bbc.co.uk

Abstract The European Broadcasting Union's AI Hub is a pioneering platform that facilitates the development and evaluation of customised AI solutions for the media industry. This collaborative ecosystem enables AI and media experts to co-create and refine artificial intelligence (AI) models designed specifically for media applications. By providing private spaces for testing AI models with proprietary content, the AI Hub guarantees comprehensive evaluations across a wide range of media content. This paper examines how these collaborative AI solutions are driving media innovation and highlights the crucial role of open source models in adapting to rapidly evolving technological landscapes. The paper will detail three key innovations: MetaRadio, which enriches radio experiences with metadata; a facial recognition system tailored for television programmes; and a Fake News Analyser designed to evaluate the reliability of news articles.

KEYWORDS: AI, generative AI, media, user-centric, evaluation, radio programme, facial recognition, disinformation

DOI: 10.69554/EVHO1458

INTRODUCTION

The field of artificial intelligence (AI) is distinguished by a rapid pace of innovation, with over 400,000 research papers published in 2023 alone. This activity is also reflected in the proliferation of AI projects on platforms such as GitHub, which currently hosts more than 340,000 AI-related projects, or Hugging Face, which has more than 120,000 projects. In such a dynamic environment, the open source model offers considerable advantages, allowing media organisations to incorporate the latest innovations and avoid the risk of being locked into proprietary technologies that could quickly become obsolete. To keep abreast of the latest open source models and adapt them to media applications, it is essential to bring together media experts and AI experts. This collaboration is fundamental to the design, testing and improvement of AI applications adapted to the specificity of media contents.

To address these needs and facilitate collaboration, the European Broadcasting Union (EBU) has created the AI Hub — a collaborative ecosystem designed to facilitate interaction between AI designers and media experts. The hub hosts a range of AI tools designed specifically for media solutions, along with user interfaces for media experts to access these tools, and application programming interfaces (APIs) for developers.

This overview describes three innovative media-oriented AI applications that have been integrated into the hub, and developed to process audio, video and text data. MetaRadio, developed by the EBU and Radio France, enables listeners of radio programmes or podcasts to browse audio content, transcripts, chapters and metadata, creating a multi-dimensional radio experience. The second application is a facial recognition system designed to handle television programmes. This system — the result of a collaboration between the EBU, the Italian public broadcasting company RAI, and RTS — aims to identify personalities in video content, ensuring that the metadata are relevant and of high quality while adhering to privacy regulations. Finally, a fake news analyser examines the linguistic properties of news articles to identify misleading writing styles and assigns a reliability score to content based on these properties.

The importance of open source AI in media

As early as 2007, recognised experts in the field of AI explained the need for and highlighted the benefits of using open

source in the development of AI; their vision is outlined by Sonnenburg *et al.*[1] Open source software allows the source code to be publicly available for anyone to view, modify and distribute, fostering collaboration, transparency and innovation. Common open source licences include the permissive MIT License, the Apache 2 License with additional patent protection, and the copyleft GNU General Public License, which requires derivative works to also be open source. The Open Responsible AI License (Open RAIL) specifically addresses AI, imposing ethical restrictions to prevent harmful applications while allowing free reuse and commercialisation except for restricted cases.

In AI, open source frameworks promote collaborative development and legal protection while ensuring responsible use. However, significant AI innovations related to foundation models, particularly in large language models (LLMs), often come from major private companies due to the substantial investment required.

The specificity of foundation models

The Stanford Institute for Human-Centered Artificial Intelligence's Center for Research on Foundation Models popularised and defined the term 'foundation model' as a model that is trained on broad data (generally using self-supervision at scale) that can be adapted (ie fine-tuned) to a wide range of downstream tasks. Foundation models such as LLMs have transformed the AI field by providing a general-purpose starting point for a wide range of applications. These models can be fine-tuned or integrated as a block in broader applications, making them highly versatile. However, because they are trained on vast amounts of data, foundation models are more difficult to evaluate than traditional models, which are intrinsically bound by their training sets. This increased complexity in their evaluation underscores the importance of developing robust and comprehensive evaluation methodologies, as understanding their performance and limitations becomes more challenging yet critical.

Avoiding the risks of closed source solutions

Relying on a closed source solution can restrict an organisation's ability to remain innovative. Embracing open source AI is a strategic move that mitigates this risk, offering media organisations the flexibility to adapt to emerging technologies. Open source AI not only ensures that media organisations can leverage the latest advancements but also fosters a culture of innovation and continuous improvement. The growing popularity and adoption of open source AI models have already significantly improved the accuracy and efficiency of media workflows. Models such as Whisper[2] for transcription, and Mixtral[3] or BERT[4] for natural language processing (NLP), developed by leading companies like OpenAI, Mistral and Google, respectively, have become go-to tools for the industry.

In the rapidly evolving field of AI, however, it is challenging to keep up with the latest developments and innovations. A hub for media applications that brings together diverse stakeholders and resources is an effective way to leverage value for the community and stay at the forefront of technology. By collaborating and sharing knowledge, media organisations can unlock new opportunities and drive growth in the industry.

A COLLABORATIVE ECOSYSTEM FOR MEDIA INNOVATION

In the fast-paced world of AI, the EBU AI Hub is dedicated to rapidly showcasing applications, providing designers with visibility and feedback to refine and enhance their tools. The AI Hub focuses not only on technology but also on inspiring creativity among AI and media experts. By sharing

their work and learning from others, participants create an environment where innovative ideas can thrive.

Enhancing AI through user-driven feedback

AI systems, inherently trained on diverse datasets, require rigorous testing to ensure they perform as expected across various scenarios. Unlike deterministic systems, AI's behaviour can be unpredictable, necessitating extensive validation to confirm its reliability and effectiveness in real-world applications. The AI Hub facilitates the use of proprietary content, providing secure private spaces and ensuring compliance with data protection regulations. Additionally, a discussion forum enhances collaboration among AI users, designers and media professionals, supporting the evaluation, refinement and responsible design of AI systems. Unlike generic platforms, the AI Hub is designed to meet the specific needs of the media community, offering custom-tailored tools for media applications that broad-market solutions do not typically address.

CUSTOM AI APPLICATIONS TO ADDRESS MEDIA SPECIFICS

Broadcasters often have specific requirements that are not fully covered by off-the-shelf open source models. To meet these needs, open source models can be customised and assembled into bespoke AI applications. This approach allows broadcasters to take advantage of the flexibility and adaptability of open source AI while ensuring that the solutions are precisely tailored to their operational needs. This paper explores three innovative tools that demonstrate the potential of tailored AI applications for media.

Enhanced radio experiences with AI

MetaRadio is an innovative application that enhances radio programmes with rich metadata, leveraging the latest advancements in transcription and diarisation technologies. This approach not only improves accessibility but also augments the user experience by integrating audio transcription and metadata into a cohesive platform.

Traditionally, radio consumption has been linear and constrained by time. However, with the integration of LLM technologies, MetaRadio transforms radio programmes, enabling enriched, structured and enhanced listening experiences. It allows users to interact with audio content, transcriptions, chapters and metadata, creating a multi-dimensional radio experience that overcomes traditional limitations. MetaRadio improves the user experience by making the consumption of radio content more interactive and dynamic. One of its key features is the ability to organise programmes into chapters with titles, summaries and keywords, making it easy for users to jump to points of interest. This non-linear exploration is particularly valuable for long programmes such as podcasts, where listeners may want to find a particular discussion or key moment without having to scroll through the entire content. For example, a listener can quickly find a particular segment, such as a guest speaker's insight or a notable event, without missing important context. In addition, the use of emoticons to annotate each speaker's turn, combined with the extraction of named entities from the text, adds another layer of interactivity, making content exploration more visual and engaging.

MetaRadio is designed to serve both casual listeners and professionals, such as journalists, who need an efficient tool for navigating and analysing radio content. The solution can be used internally by broadcasters or deployed on public platforms, offering audiences a new and enriched way to consume radio content.

As shown in Figure 1 (top), MetaRadio organises content into chapters, providing users with a structured overview of the programme and enabling seamless navigation. These chapters, consisting of a title, summary and keywords, follow the chronology of the programme, maintaining the structured flow of the radio broadcast. However, MetaRadio

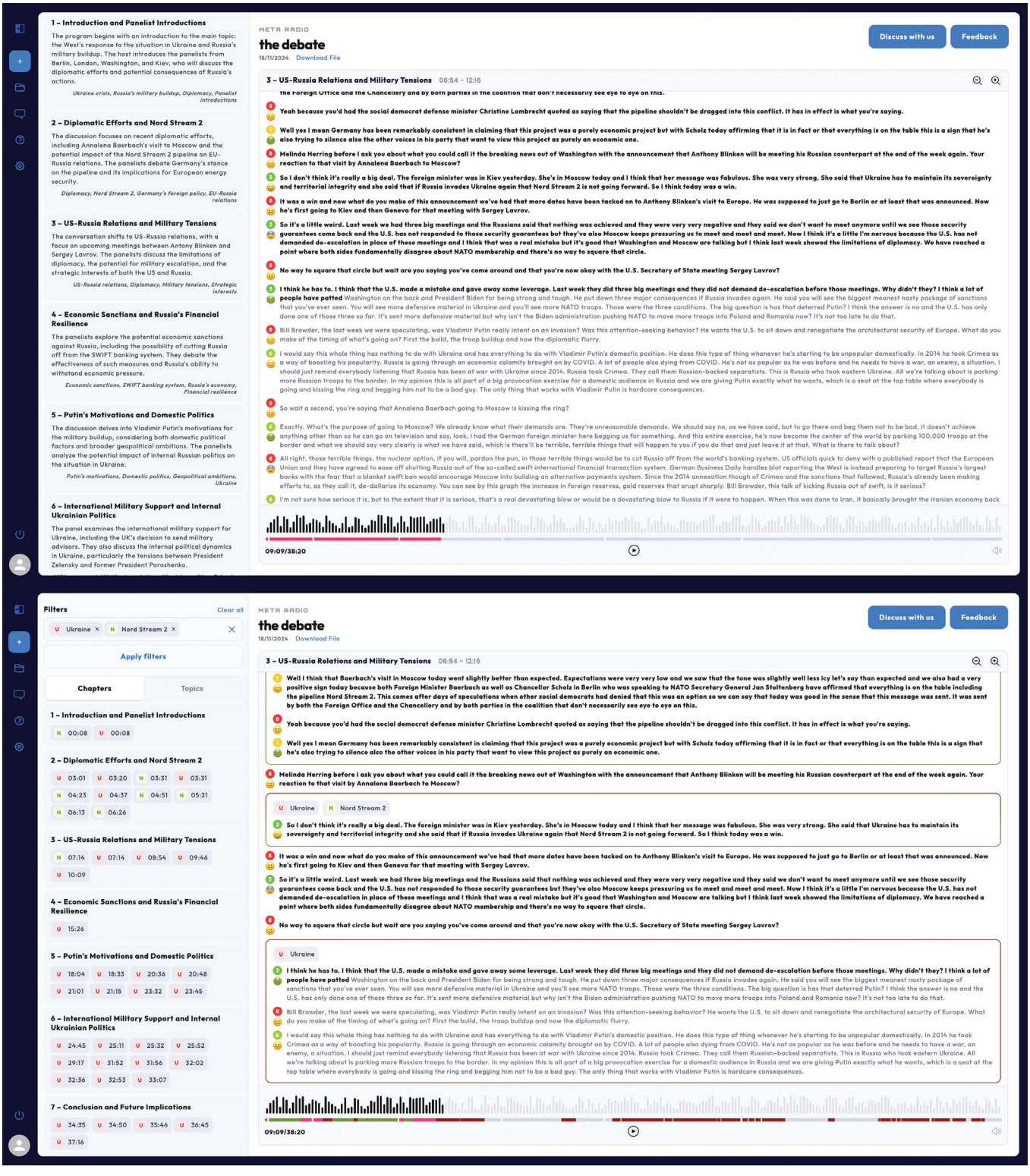

Figure 1 MetaRadio: A multi-dimensional radio experience featuring linear navigation through chapters (top) and non-linear navigation via topics (bottom)

also supports non-linear navigation, as shown in the lower part of the figure, where users can jump directly to specific named entities and topics. This dual approach provides flexibility, enabling both immersive, step-by-step exploration and efficient, targeted searching.

Built on open-source AI models

MetaRadio integrates state-of-the-art open source AI tools to improve the processing of radio content. For transcription, we use Faster-Whisper, an optimised version of OpenAI's Whisper model. Faster-Whisper uses CTranslate2 — a high-performance inference engine for Transformer models — which significantly reduces transcription time, making it ideal for large-scale audio processing tasks. For content annotation, MetaRadio uses fine-tuned versions of BERT (Bidirectional Encoder Representations from Transformers), the open source language model developed by Google. BERT's contextual understanding and ability to generate dynamic word embeddings enable accurate semantic tagging of transcriptions. This approach outperforms traditional static word-embedding models because BERT takes into account the context of each word occurrence, improving the accuracy of tasks such as named entity recognition.

To generate chapters and structure radio content, MetaRadio uses Mixtral 8x7B — an open-weight LLM developed by Mistral AI. The mixture of experts architecture used by Mixtral 8x7B enables it to handle complex linguistic tasks efficiently while maintaining coherence over extended contexts. This capability makes it particularly well suited to creating meaningful chapters in radio programmes. As an open-weight model, Mixtral can be fine-tuned and hosted within the hub, offering flexibility and customisation to suit specific content requirements, while also allowing control over the versioning of customised models.

MetaRadio demonstrates the transformative potential of language models and AI to innovate traditional podcast formats. By providing broadcasters with advanced strategies for audience engagement and content accessibility, this project promotes a more flexible, personalised and interactive radio consumption experience. MetaRadio also addresses the needs of journalists and documentarians by providing efficient navigation through radio content, thereby improving productivity and content management within the broadcast industry. It also facilitates the reuse of existing programmes, creating new value from previously broadcast content.

The Fake News Analyser

The Fake News Analyser is an innovative tool designed to combat the growing threat of misinformation in modern media. With the rise of online news platforms and social media, misleading content has become more prevalent, often bypassing traditional editorial filters. The Fake News Analyser uses advanced NLP techniques to detect and flag potentially deceptive information in news articles. It examines the linguistic properties of the content, focusing on lexical, categorical and semantic features to identify patterns commonly associated with deceptive writing styles.

This approach is based on the assumption that fake news uses specific writing strategies to manipulate readers. For example, fake news articles may use exaggerated language, sensationalism or a lack of sound reasoning to obscure the truth and mislead the audience. A key feature of fake news is that it relies on emotional appeal rather than logical persuasion, omitting well-structured arguments to make it harder for readers to evaluate the information critically. This increases the likelihood of misinformation spreading unchecked.

The system assigns each article a reliability score based on the presence of these linguistic markers. Rather than making a binary decision about whether an article is definitively true or false, the Fake News Analyser quantifies the reliability of content by providing a gradient score that reflects the likelihood of trustworthiness. This provides a fast, automated

way to flag suspicious content without the need for immediate fact-checking.

As shown in Figure 2, the analyser visually highlights key textual elements that contribute to the article's credibility score. Text that negatively impacts the score is highlighted in red, while features that enhance credibility are highlighted in green. This visual representation allows users to quickly identify problematic sections and those that strengthen the article's credibility. The aim is not to replace human judgment, but rather to complement it with data-driven insights into the linguistic integrity of content.

Linguistic features and categorisation

As illustrated in Figure 3, the Fake News Analyser uses a comprehensive model that groups linguistic features into three key categories:

- *Lexical features:* This involves analysing the individual words used in the article. Fake news often uses specific word choices that may appear sensational or manipulative, such as excessive use of emotionally charged terms designed to provoke strong reactions. The analyser evaluates the frequency and context of such words to determine whether they contribute to misleading narratives.
- *Categorical features:* Categorical analysis uses a dictionary-based approach in which words are grouped into predefined categories. These categories capture broader emotional and perceptual aspects of language, such as emotion (eg anger, fear), perception (eg seeing, believing) or cognition (eg thinking, knowing). By identifying the overuse or misuse of these categories, the Fake News Analyser can detect when content is appealing to emotions or manipulating cognitive biases rather than presenting balanced, fact-based information. Fake news articles often use these categories to sway readers' opinions through emotional triggers rather than objective facts.

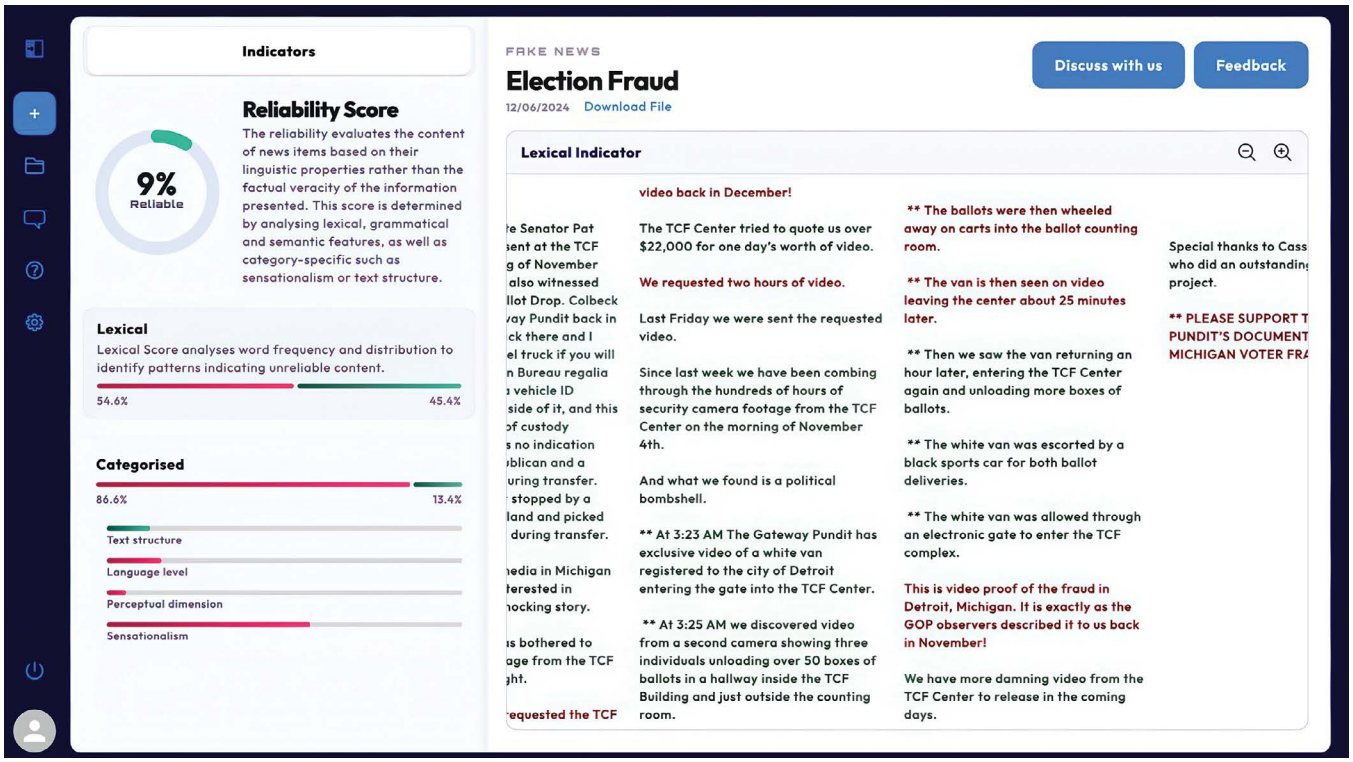

Figure 2 The Fake News Analyser, computing reliability scores by highlighting linguistic properties

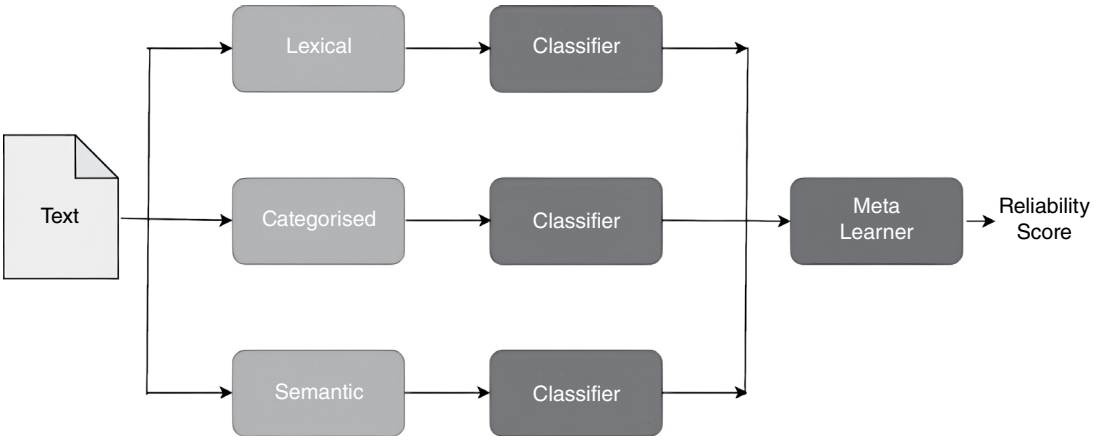

Figure 3 Architecture of the Fake News Analyser — Lexical, categorical and semantic features are processed by base classifiers, which provide scores that the meta learner combines to calculate the reliability score

- *Semantic features:* Using NLP models such as BERT (Bidirectional Encoder Representations from Transformers), the Fake News Analyser embeds text in a high-dimensional space to capture deeper contextual representations of words. This enables the detection of more complex forms of deception that may not rely on overtly sensational language, but still aim to mislead through more subtle tactics.

Each of these categories is processed by base classifiers, which independently assess the probability that content within their respective domains (lexical, categorical or semantic) could be deceptive. These probabilities are then combined by a meta-learner — a higher-level model designed to aggregate the insights from the base classifiers into a final reliability score. This hierarchical architecture enables the analyser not only to detect fake news more accurately, but also to provide an explainable breakdown of the decision-making process by quantifying the influence of each feature type.

Reliability score and explainability

One of the key strengths of the Fake News Analyser is its emphasis on explainability. While many AI systems operate as 'black boxes', delivering results without transparency, the Fake News Analyser provides clear insight into how results are derived. As shown in Figure 3, users can trace the impact of each feature category (lexical, categorical and semantic) on the final reliability score.

For example, if an article's lexical score is low, users can investigate the specific words triggering the reduction. If the categorical score raises concerns, the analyser can highlight emotional or perceptual categories that have been overused, indicating possible attempts at manipulation. This level of transparency allows news organisations to quickly identify and address problematic content, ensuring the integrity of their reporting.

In addition, the system's ability to colour-code linguistic properties (red for negative influence, green for positive influence) makes it easy to visualise areas of concern within an article. This not only aids in the detection of fake news, but also serves as a valuable tool for content review and editorial decision-making.

Applications and scalability

The Fake News Analyser has a wide range of applications in the media industry, particularly for news organisations that process large volumes of content on a daily basis. By providing a scalable solution capable of analysing large volumes of text automatically,

the analyser helps journalists efficiently prioritise content flagged for low reliability scores. Traditionally, journalists rely on experience and intuition to identify misleading content. However, the Fake News Analyser uses sophisticated statistical rules that are often difficult to articulate, augmenting journalistic expertise with machine-driven insights.

The analyser can be seamlessly integrated into existing workflows for internal content verification, or deployed on public platforms to alert readers and encourage further investigation of flagged content. By providing a tool that combines scalability with deep linguistic analysis, the Fake News Analyser enhances news organisations' ability to maintain high standards of accuracy and integrity, even in the face of rapidly evolving misinformation.

A facial recognition system tailored for television programmes

In the field of object recognition, closed-set and open-set settings are fundamental approaches that address different challenges in identifying and categorising objects. In the open-set setting, an AI model must not only discriminate between the classes of the training set or more generally the set of classes to be identified but also determine whether a class has not yet been encountered. Conversely, in the closed-set setting, a model is tasked with recognising a fixed set of classes that remains consistent throughout both the training and testing phases. Both closed and open sets can involve a zero-shot learning approach, where the model recognises categories that it has not explicitly seen during training but which are part of the set of classes to be identified. In the context of facial recognition, identifying individuals in an open set is particularly relevant to the media, as television programmes and films often introduce both recurring characters and new, unpredictable ones. Consequently, in practical applications, the facial recognition system must constantly adapt and decide whether faces represent new individuals that should be added to the database.

In summary, open-set identification in the context of facial recognition for media must accommodate the constant evolution of media content and ensure that systems are both effective in recognition and able to identify unknown individuals who should be recognised. To establish the terminology in the context of facial recognition, the reference database comprises two primary elements: the thesaurus and the gallery. The thesaurus contains the names or identifiers of individuals to be recognised, while the gallery consists of a collection of images for each individual.

In general, facial recognition algorithms compute one-to-many similarity scores to determine the specific identity of a person in a probe image. In the context of facial recognition, the individual in the probe image may or may not be previously known to the system. If the probe face does not meet the acceptance criterion, such as having a similarity score below a predefined threshold, the person is considered unknown. In such cases, subject to legal and ethical considerations, the system may be enriched with the new individual's information, which would be added to both the thesaurus and the gallery.

The classical machine-learning performance metrics for facial recognition on open-set are detailed by Myllyaho *et al.*[5] While these metrics provide valuable insights into absolute performance, they do not account for the practical utility of the tool. The user-centric metrics that we developed therefore focus on the specific needs of documentalists. Furthermore, a comprehensive evaluation of facial recognition technology must consider a broader spectrum of factors, including computational complexity and adherence to legal standards such as the EU General Data Protection Regulation (GDPR). Our approach simplifies the implementation by focusing on key operational and legal considerations, ensuring that the deployment of facial recognition technologies is both effective and compliant with privacy regulations.

User-centric metrics

User-centric metrics are designed to meet the specific needs of documentalists who are responsible for managing annotations and efficiently identifying relevant content. Documentalists perform two critical functions: annotation management and relevant content retrieval. During the annotation process, documentalists are responsible for verifying and naming individuals who should be added to the thesaurus. This role requires ensuring that the system prioritises individuals who contribute actively to the programme's content over random bystanders or background characters, while also ensuring compliance with GDPR for lawful inclusion in the database.

During the search phase, documentalists need access to precisely annotated intervals rather than continuous frame-by-frame data that can overwhelm the system with irrelevant information. This flood of data not only obscures the search process but also increases the likelihood of errors requiring further post-processing to eliminate false automatic annotations. In our system, an interval is defined by two criteria: it must be less than 30 seconds in duration and contain a maximum of three distinct active individuals. This interval-based annotation approach ensures that the dataset is aligned with the operational needs of documentalists, emphasising the relevance of the recognised faces without compromising the efficiency of the system.

To adapt the performance metrics to the specific needs of the documentalists for annotation and gallery management, we introduce gallery candidate precision (GCP). GCP measures the precision with which a facial recognition system identifies and suggests candidates for inclusion in the reference database. Optimising GCP significantly reduces the workload of documentalists in managing the gallery.

In conclusion, user-centric metrics in facial recognition for media applications must prioritise the specific operational needs of documentalists. Key metrics such as GCP and the strategic use of interval-based annotations are critical to making these systems practical, effective and user-centric.

Facial recognition pipeline for video

This section discusses how the algorithms developed by the EBU optimise user-centric performance metrics. Figure 4 illustrates the process of facial recognition on video and highlights the face selection block developed by the EBU to optimise user-centric performance metrics. The

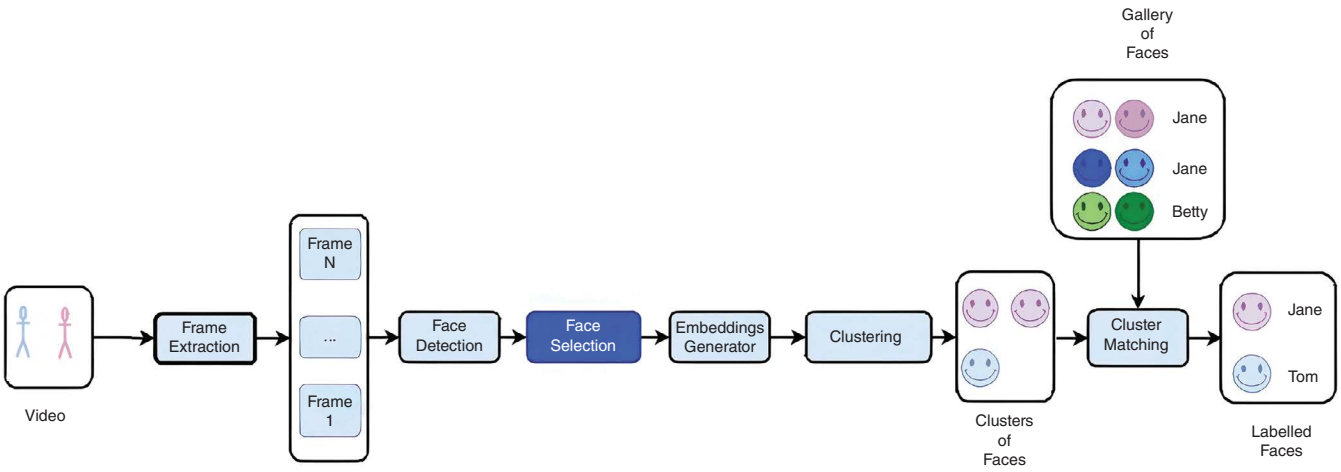

Figure 4 Facial recognition processing blocks for video

classical facial recognition processing blocks are detailed in Du *et al.*[6] and Mendes *et al.*,[7] starting with frame extraction, the transformation of detected faces into high-dimensional vectors using deep neural networks, and the subsequent clustering and matching process to associate these faces with individuals in the gallery. The EBU's face selection block leverages prior knowledge of the shooting style to filter out faces that are detected but do not belong to active personalities in the show.

Leveraging filming style for efficient face selection

In television programmes and films, the style of the cameraman has a significant impact on the essence of the production. The shooting style is designed to emphasise the importance of key characters or personalities, making them visually dominant on the screen. These key characters are the active participants in the programme who should be present in the metadata. The processing pipeline shown in Figure 4 uses the face selection block to identify relevant faces based on several factors. Figure 5 illustrates this process in a television show.

Impact of the face selection algorithms on user-centric performance metrics

The face selection algorithm aims to optimise user-centric metrics, enhance identification performance and reduce complexity. Figure 6 illustrates the effectiveness of the algorithm in a television show where the public is often present in the background of the guests and hosts.

We have observed that commercial solutions with large galleries can lead to a higher probability of individuals in the audience of the show resembling someone already present in the gallery. This increased likelihood of similarity can lead to misidentification and the attribution of names in the thesaurus to anonymous individuals.

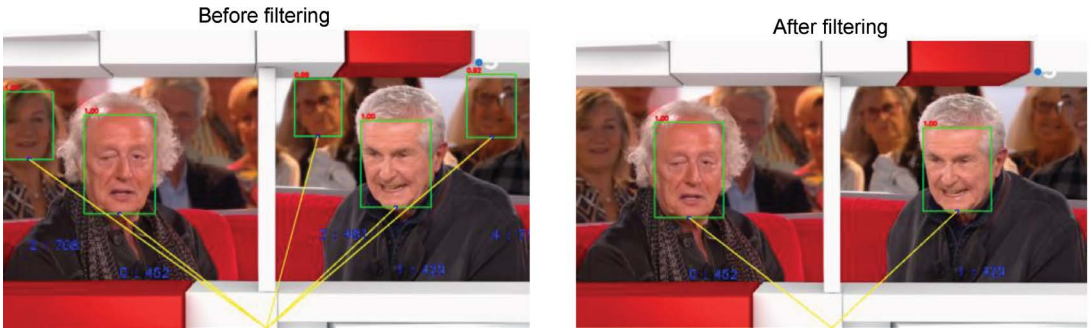

Figure 5 Face selection processing before and after face selection

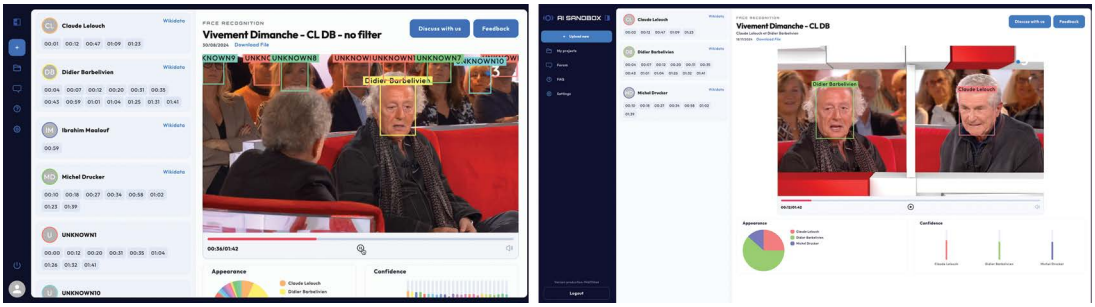

Figure 6 Effect of the face selection in a television show

Facial recognition and GDPR compliance

Under privacy legislation such as the GDPR, the use of facial recognition technology on spectators or inactive participants at public events like television programmes raises privacy concerns. While the GDPR allows limited exemptions for journalism, this does not permit unrestricted biometric processing without safeguards. To avoid violating privacy legislation, the processing of biometric data for identification purposes therefore requires a data protection impact assessment.

The development of algorithms to detect and filter out the facial data of non-participants' can help with GDPR compliance. For active participants who consent or are covered by journalistic exceptions, the processing adheres to data minimisation and aligns with purpose limitation, and constitutes data protection by design. This approach enables facial recognition for television content while preventing GDPR violations related to consent and individual rights. For non-participating celebrities, the journalistic exemption should be applied restrictively based on legitimate public interest rather than curiosity about private life. This method prioritises privacy protection while allowing necessary facial recognition functionality.

CONCLUSION

The integration of open source AI models into media applications represents a transformative shift in the broadcast industry. By harnessing the adaptability and collaborative potential of open source technologies, media organisations can improve workflows, increase content accessibility and maintain a competitive edge in a rapidly evolving landscape. The EBU AI Hub exemplifies this approach by providing a collaborative ecosystem where AI and media experts can co-create, test and refine AI models tailored to the unique needs of the media sector. Key innovations such as MetaRadio, the Fake News Analyser and the facial recognition system for television programmes demonstrate the practical benefits of open source AI. By fostering an environment where media professionals and AI experts can share knowledge and resources, the EBU AI Hub not only accelerates innovation but also ensures that AI solutions are practical, effective and aligned with the operational needs of the media industry.

References

(1) Sonnenburg, S., Braun, M. L., Ong, C. S., Bengio, S., Bottou, L., Holmes, G., LeCunn, Y., Muller, K. R., Pereira, F., Rasmussen, C. E. and Ratsch, G. (2007) 'The Need for Open Source Software in Machine Learning', *Journal of Machine Learning Research*, Vol. 8, pp. 2443–66.

(2) Radford, A., Kim, J. W., Xu, T., Brockman, G., McLeavey, C. and Sutskever, I. (2023) 'Robust Speech Recognition via Large-Scale Weak Supervision', in: Proceedings of the International Conference on Machine Learning, Hawaii, 3rd July, pp. 28492–518.

(3) Jiang, A. Q., Sablayrolles, A., Roux, A., Mensch, A., Savary, B., Bamford, C., Chaplot, D. S., Casas, D. D., Hanna, E. B., Bressand, F. and Lengyel, G. (2024) 'Mixtral of Experts', arXiv preprint arXiv:2401.04088.

(4) Devlin, J., Chang, M. W., Lee, K. and Toutanova, K. (2018) 'BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding', arXiv preprint arXiv:1810.04805.

(5) Myllyaho, L., Raatikainen, M., Männistö, T., Mikkonen, T. and Nurminen, J. K. (2021) 'Systematic Literature Review of Validation Methods for AI Systems', *Journal of Systems and Software*, Vol. 181, Article 111050.

(6) Du, H., Shi, H., Zeng, D., Zhang, X. P. and Mei, T. (2022) 'The Elements of End-to-End Deep Face Recognition: A Survey of Recent Advances', *ACM Computing Surveys*, Vol. 54, No. 10(Suppl.), pp. 1–42.

(7) Mendes, P. R., Busson, A. J., Colcher, S., Schwabe, D., Guedes, Á. L. and Laufer, C. (2020) 'A Cluster-Matching-Based Method for Video Face Recognition', in: Proceedings of the Brazilian Symposium on Multimedia and the Web, São Luís, 30th November, pp. 97–104.